

Proyecto Fin de Grado

Ingeniería de la Energía

Aplicación y validación de la técnica de empaquetamiento de días a la base radiométrica del GTER

Autor: Álvaro Machuca Calzado

Tutor: Manuel Antonio Silva Pérez

Dep. Ingeniería Energética
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2016



CONTENIDO

1	Introducción	3
2	Conceptos previos	5
3	Objetivos	12
4	Método de empaquetamiento de días	13
5	Aplicación del método.....	16
5.1	Cálculo de la envolvente de cielo despejado	16
5.2	análisis de similitud	23
6	Aplicación a la base de datos GTER.....	24
6.1	Problemática del método.....	25
6.2	Aplicación del método a datos horarios	27
6.2.1	Ampliación minutal de datos horarios	29
6.3	Empaquetamiento de días nublados	32
6.4	Generación de años sintéticos	35
6.4.1	Generación año 2014	35
7	Análisis estadístico de modelos	44
7.1	Análisis KSI (%).....	47
7.2	Otros indicadores estadísticos	49
7.2.1	Indicadores de dispersión	49
7.2.2	Indicadores de rendimiento	50
8	Conclusiones.....	53
9	Referencias	54
10	ANEXO I	55
11	ANEXO II	58

ÍNDICE DE FIGURAS

FIGURA 1. IRRADIANCIA DÍA CIELO DESPEJADO. [6]	6
FIGURA 2. IRRADIANCIA CIELO NO DESPEJADO. DNI (ROJO) IRRADIANCIA GLOBAL (AZUL).....	7
FIGURA 3. COORDENADAS CELESTES HORIZONTALES.	8
FIGURA 4. ALTURA SOLAR A LO LARGO DEL DÍA 187.	9
FIGURA 5. VALORES DNI ESCALA HORARIA.	10
FIGURA 6. VALORES DNI ESCALA DIEZ MINUTAL.	10
FIGURA 7. VALORES DNI ESCALA MINUTAL.....	11
FIGURA 8. CURVA DATOS DNI MINUTALES CON CURVA DE DÍA DESPEJADO Y DÍA ADIMENSIONALIZADO.....	15
FIGURA 9. CURVA DATOS MINUTALES DNI, 22 DE AGOSTO 2010.....	16
FIGURA 10. CURVA DATOS MINUTALES DNI CON PUNTOS DE CIELO DESPEJADO, 22 DE AGOSTO 2010.	18
FIGURA 11. CURVA PUNTOS CIELO DESPEJADO.	19
FIGURA 12. AJUSTE LINEAL ROBUSTO.	20
FIGURA 13. AJUSTE LINEAL ROBUSTO DE CIELO DESPEJADO.	21
FIGURA 14. DATOS MINUTALES DNI Y ENVOLVENTE DE CIELO DESPEJADO.	21
FIGURA 15. DÍA 22 DE AGOSTO ADIMENSIONALIZADO.	22
FIGURA 16. DÍA 22 DE AGOSTO ADIMENSIONALIZADO Y APROXIMACIÓN DE 1000 PUNTOS.	23
FIGURA 17. APLICACIÓN MÉTODO EMPAQUETAMIENTO. 7 DE ENERO.	24
FIGURA 18. APLICACIÓN MÉTODO EMPAQUETAMIENTO. 2 DE ENERO.	25
FIGURA 19. APLICACIÓN MÉTODO EMPAQUETAMIENTO. 4 DE ENERO.	25
FIGURA 20. APLICACIÓN MÉTODO EMPAQUETAMIENTO DATOS HORARIOS. 7 DE ENERO.	27
FIGURA 21. APLICACIÓN MÉTODO EMPAQUETAMIENTO DATOS HORARIOS. 8 DE ENERO.	28
FIGURA 22. APLICACIÓN MÉTODO EMPAQUETAMIENTO DATOS HORARIOS. 4 DE ENERO.	28
FIGURA 23. APLICACIÓN MÉTODO EMPAQUETAMIENTO DATOS HORARIOS. 22 DE DICIEMBRE (SUPERIOR). 5 DE ENERO(INFERIOR).	29
FIGURA 24. APLICACIÓN MÉTODO EMPAQUETAMIENTO DATOS HORARIOS CON AMPLIACIÓN MINUTAL. 22 DE DICIEMBRE (SUPERIOR). 5 DE ENERO(INFERIOR).	30
FIGURA 25. CÁLCULO DE ENVOLVENTE DÍA 8 DE ENERO. APLICACIÓN SIN AMPLIACIÓN HORARIA (ARRIBA). APLICACIÓN CON AMPLIACIÓN HORARIA (ABAJO).	31
FIGURA 26. VALORES MÁXIMOS LN(DNI) PARA LA SERIE ANUAL 2014.	33
FIGURA 27. AJUSTE LINEAL ROBUSTO ENVOLVENTE GLOBAL.	34
FIGURA 28. AJUSTE LINEAL ROBUSTO ENVOLVENTE GLOBAL DE MÁXIMO VALOR.	34
FIGURA 29. COMPARACIÓN VALOR ACUMULADO MENSUAL.	36
FIGURA 30. COMPARACIÓN VALOR ACUMULADO DIARIO.....	37
FIGURA 31. ERROR DIARIO AÑO SINTÉTICO RESPECTO A AÑO MEDIDO.....	37
FIGURA 32. DÍA 2 ENERO DE 2014 DESEMPAQUETADO CON DÍA 327 DE 2000. ERROR 143%.	38
FIGURA 33. DÍA 8 FEBRERO DE 2014 DESEMPAQUETADO CON DÍA 66 DE 2001. ERROR 151%.	38
FIGURA 34. DÍA 12 FEBRERO DE 2014 DESEMPAQUETADO CON DÍA 340 DE 2003. ERROR 266%.	39
FIGURA 35. DÍA 13 DICIEMBRE DE 2014 DESEMPAQUETADO CON DÍA 67 DE 2011. ERROR 128%.	39
FIGURA 36. ERROR DIARIO AÑO SINTÉTICO RESPECTO A AÑO MEDIDO. SIN ERRORES EXTREMOS MAYORES DE 100%.....	40
FIGURA 37. DÍA 10 ENERO DE 2014 DESEMPAQUETADO CON DÍA 34 DE 2002. ERROR 60%.	41
FIGURA 38. DÍA 21 ENERO DE 2014 DESEMPAQUETADO CON DÍA 104 DE 2000. ERROR 59%.	41
FIGURA 39. DÍA 20 FEBRERO DE 2014 DESEMPAQUETADO CON DÍA 61 DE 2002. ERROR 63%.	42
FIGURA 40. DÍA 28 MARZO DE 2014 DESEMPAQUETADO CON DÍA 324 DE 2011. ERROR 77%.	42
FIGURA 41. ERROR DIARIO AÑO SINTÉTICO RESPECTO A AÑO MEDIDO. SIN ERRORES EXTREMOS MAYORES DE 50%.....	43
FIGURA 42. COMPARACIÓN DATOS ACUMULADOS ANUALES ENTRE AÑOS GENERADOS 2014.....	44
FIGURA 43. LÍMITE TOLERABLE VALOR DN. MÉTODO LARRAÑETA (<i>Ibndet</i>) Y MÉTODO 1 (<i>Ibnsynth</i>).....	47
FIGURA 44. FUNCIÓN DE DISTRIBUCIÓN. MÉTODO LARRAÑETA (<i>Ibndet</i>) Y MÉTODO 1 (<i>Ibnsynth</i>).....	47

1 INTRODUCCIÓN

En las últimas décadas, ha aumentado la preocupación por el medio ambiente, el cual pelagra en gran parte por el uso de combustibles fósiles, los cuales tienen asociadas unas emisiones de gases de efecto invernadero que han agravado el llamado calentamiento global.

Pero esto no ha sido lo único que hay en contra de los combustibles fósiles, las dos crisis del petróleo, los sucesivos conflictos en los países productores, la no distribución natural de estos recursos, los cuales se encuentran en determinados lugares de la tierra, han llevado a un nuevo pensamiento, el de buscar otras fuentes de energía, que sean accesibles para todos los países, que estén relacionadas a unas bajas o nulas emisiones y que sean de fuente renovable.

El concepto de energía renovable se refiere a aquella fuente de energía que es virtualmente inagotable, esto es, que se renueva por medios naturales o que de por sí contiene una fuente inmensa de energía.

En los últimos años, se ha realizado una gran inversión en la tecnología encargada de explotar estas fuentes de energía inagotables. Muchas de ellas, han sido utilizadas por el hombre desde hace siglos, cabe destacar la energía eólica, que, aunque hasta hace unos años no se ha utilizado para la generación de electricidad, sí ha sido utilizada como fuente de energía mecánica.

Este tipo de energía supone, en algunos casos, un alto coste tanto de inversión como de mantenimiento, es por ello, que la gran mayoría de países que invierten en ellas suelen ser países desarrollados y que tienen una alta dependencia energética, ya puede ser del petróleo, del gas o de otros tipos de hidrocarburos.

A pesar de ello, también vemos altas inversiones en estas tecnologías por países con grandes reservas de petróleo y derivados, como es el caso de Estados Unidos, que, a pesar de ser productor de gas y petróleo, su interés por el medio ambiente ha hecho que se apueste por estas nuevas tecnologías.

En el caso de la Unión Europea, tratados como “Europa 2020”, buscan aumentar la participación de las energías renovables en la generación de energía de cada país perteneciente a la Unión.

Aunque los objetivos de estos tratados no lleguen a alcanzarse, esta clase de propuestas muestran un interés por parte de los países y de las instituciones de apostar por estas tecnologías.

En el caso de España, la energía renovable ha estado presente a lo largo del siglo XX, principalmente en forma de energía hidráulica, ha sido en la última década de este siglo, cuando los gobiernos han comenzado a impulsar de manera importante otro tipo de energía de esta naturaleza.

En España, podemos destacar en los últimos años la apuesta por la energía eólica y solar.

Otras fuentes de energía de carácter renovable son la biomasa, también utilizada por el hombre desde hace siglos, la energía hidráulica o la energía solar, como hemos nombrado en el caso de España.

Cuando hablamos de energía solar nos referimos a aquella procedente del sol, que llega a la tierra en forma de radiación. El aprovechamiento de esta energía se produce mediante paneles que o bien transforman directamente esta radiación solar en energía eléctrica, tecnología fotovoltaica, o bien utilizan esta energía para calentar un fluido térmico mediante la concentración, hablamos de tecnología de concentración solar.

Dicho esto, conociendo que la energía eléctrica debe ser consumida en el momento en el que se produce, y viendo que la fuente de esta producción tiene como origen el sol, es importante conocer qué energía aporta el sol, durante cuánto tiempo y en qué medida, por ello a continuación, introduciremos conocimientos previos sobre la energía que procede de esta fuente.

2 CONCEPTOS PREVIOS

La energía procedente del sol es considerada la más limpia y abundante de la que dispone la tierra, y por ello, los seres vivos. Sin embargo, se trata de una fuente discontinua y no controlable. Además, no incide de igual modo en todas las zonas de la tierra, por lo que su aprovechamiento y utilización es limitado y complejo.

La energía solar, antes de incidir sobre la tierra sufre un proceso de atenuación debido a su paso por la atmósfera. Esta atenuación no es la misma en todas las partes de la tierra, y depende de muchos factores, que están en constante cambio. Por ejemplo, el paso de nubes, sustancias en suspensión en la atmósfera, o incluso la época del año, ya que el movimiento de la tierra respecto a sí mismo y respecto al sol, influye en gran medida sobre la energía que llega procedente del sol.

De esos elementos, uno de los más influyentes es el paso de nubes. El hecho de que existan nubes en el cielo puede producir que en una zona en la que suele llegar un alto valor de radiación, termine por llegar un valor nulo de la misma. Pero la existencia de una nube no tiene como único resultado la llegada de radiación nula. Existen muchos tipos de nube, y de igual forma, estas se pueden encontrar a distinta altura en la atmósfera. Todo esto va a influir en gran medida en la radiación resultante.

No hace falta decir, que como el sol es la fuente de energía, la falta del mismo supone un aporte energético nulo, esto es, la energía solar es una fuente que además de discontinua y dependiente de la atmósfera del lugar, sólo está disponible unas ciertas horas del día, y es de carácter estacional, ya que, dependiendo de la estación del año, contamos con más horas de sol que en otras estaciones. Esto hace que los valores de la irradiancia no se mantengan constantes todo el año, y que estos difieran de un día para otro.

La irradiancia, es la potencia de la radiación que incide en un instante sobre una superficie determinada, y es medida en W/m^2 .

Esta irradiancia la podemos clasificar de la siguiente forma:

- Irradiancia directa: es la que llega a un determinado lugar procedente directamente del disco solar. Cuando es medida sobre un plano perpendicular a la dirección de los rayos del sol, hablamos de irradiancia directa normal (DNI, Direct Normal Irradiance).
- Irradiancia difusa: procedente de toda la bóveda celeste excepto del disco solar. Supone un pequeño porcentaje respecto a la irradiancia directa, pero esto ocurre cuando hablamos de un día despejado. Cuando se trata de un día nublado, la irradiancia difusa, es la mayoritaria, ya que las nubes evitan que la irradiancia directamente procedente del disco solar, llegue a la tierra.
- Irradiancia global: puede ser entendido como la suma de la radiación directa y la difusa sobre una misma superficie. Es el total de la radiación que llega a un lugar determinado.

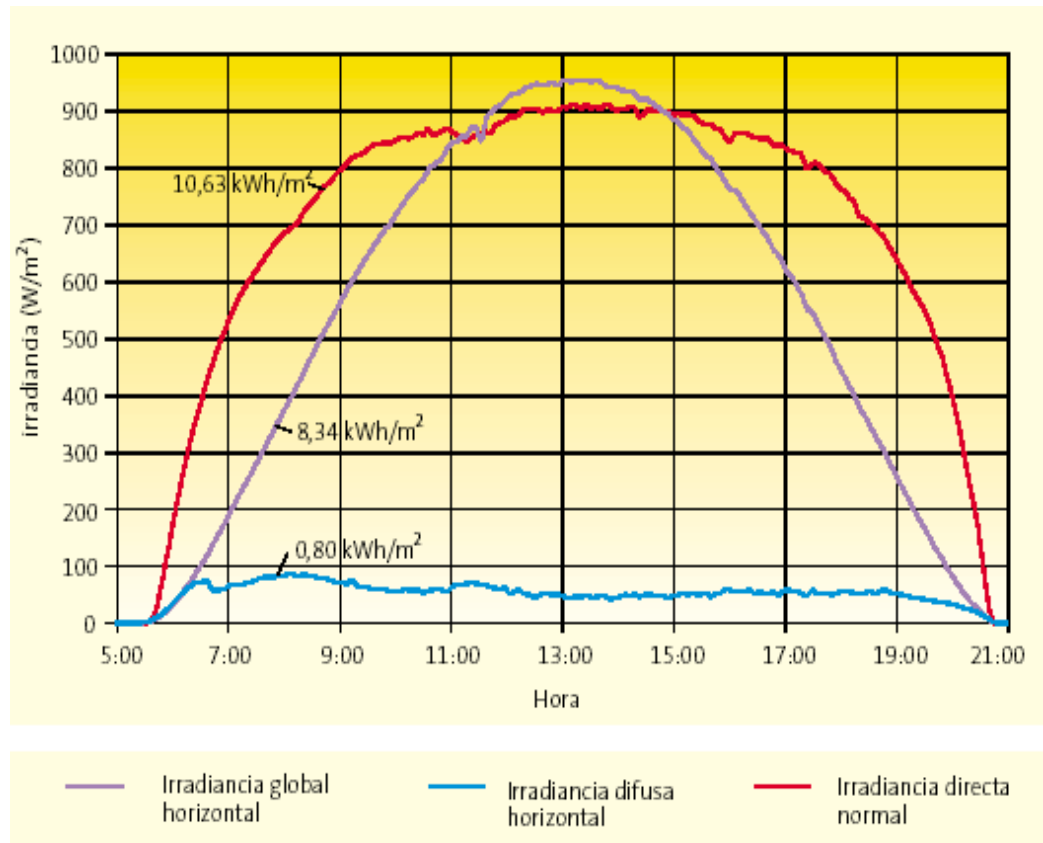


Figura 1. Irradiancia día cielo despejado. [6]

En la figura 1, vemos representadas las distintas irradiancias para un día de cielo despejado, en el que no ha habido un paso de nubes importantes o que hayan interferido en la medida. Sin embargo, la forma de estas curvas cambia en gran medida si el día ha sido un día nublado.

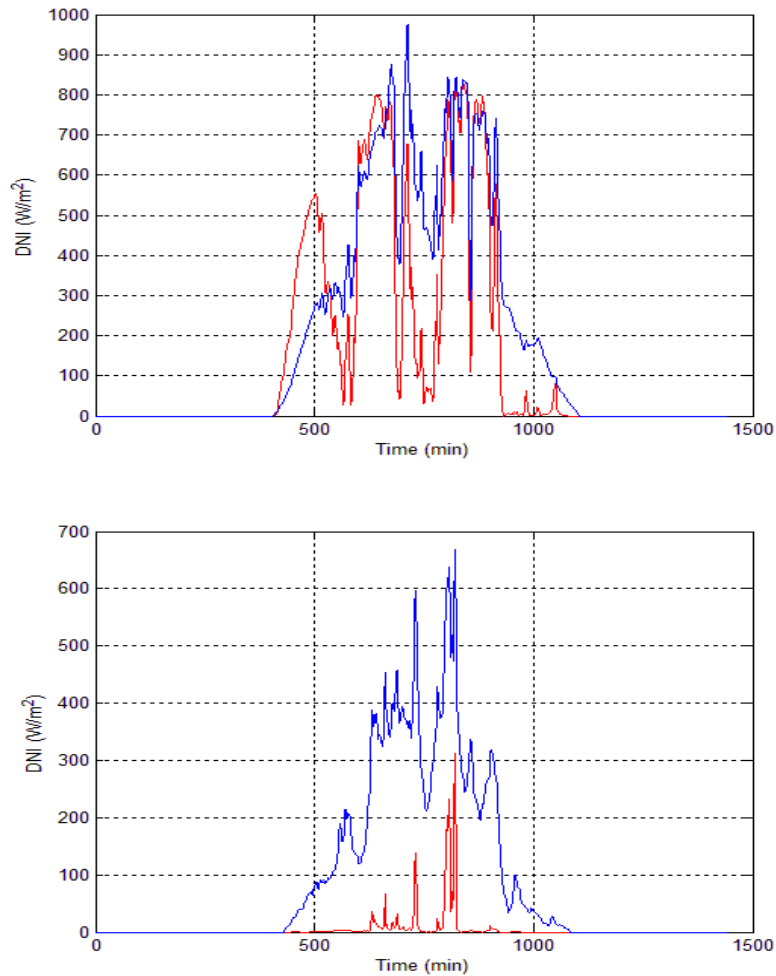


Figura 2. Irradiancia cielo no despejado. DNI (Rojo) irradiancia global (Azul).

En la figura 2 podemos observar dos días con nubosidad. Estos, a pesar de ser nublados tienen una forma muy distinta, en la imagen superior podemos hablar de un día en el que el aparato que ha medido ha sufrido intermitencias de nubes en las horas centrales del día

Mientras que en la segunda imagen de la figura 2, el día corresponde a un día nublado en el que las nubes han interferido sobre la medición durante la mayor parte del día, por lo que no hablamos de nubes pasantes, sino de cielo nublado durante todo el tiempo de medida.

Otro concepto importante es la altura solar:

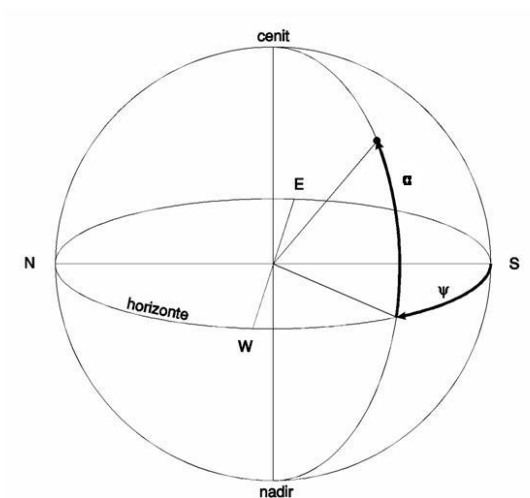


Figura 3. Coordenadas celestes horizontales.

La posición de un astro, en astronomía se determina mediante coordenadas esféricas. Aunque la distancia a estos astros sea desconocida, por lo que se considera su posición proyectada sobre una esfera de radio unidad y de centro arbitrario, denominada esfera celeste.

Las coordenadas celestes horizontales de un astro, como se muestra en la figura 3, son el acimut (ψ) y la altura (α).

El acimut, arco en el horizonte celeste que comprende entre el punto cardinal Sur y el punto donde el círculo vertical que pasa por el astro corta al horizonte. Se cuenta a partir del Sur, de 0° a $\pm 180^\circ$, positivamente hacia el Oeste y negativamente hacia el Este.

La altura, se trataría del arco de dicho círculo vertical comprendido entre el astro y el horizonte. Se mide a partir del horizonte de 0° a 90° , positivamente hacia el cenit y negativamente hacia el nadir.

Para conocer la posición del sol, se utiliza el ángulo cenital. Este ángulo, toma como referencia el plano vertical que pasa por el cenit para definir la posición del sol en un momento dado. En nuestro caso, para aplicar el método de Fernández Peruchena [1], es necesario aplicar unos filtros que utilizan el ángulo complementario al ángulo cenital, conocido como la altura solar. La altura solar, toma entonces como referencia el plano horizontal, al contrario que hacía el ángulo cenital.

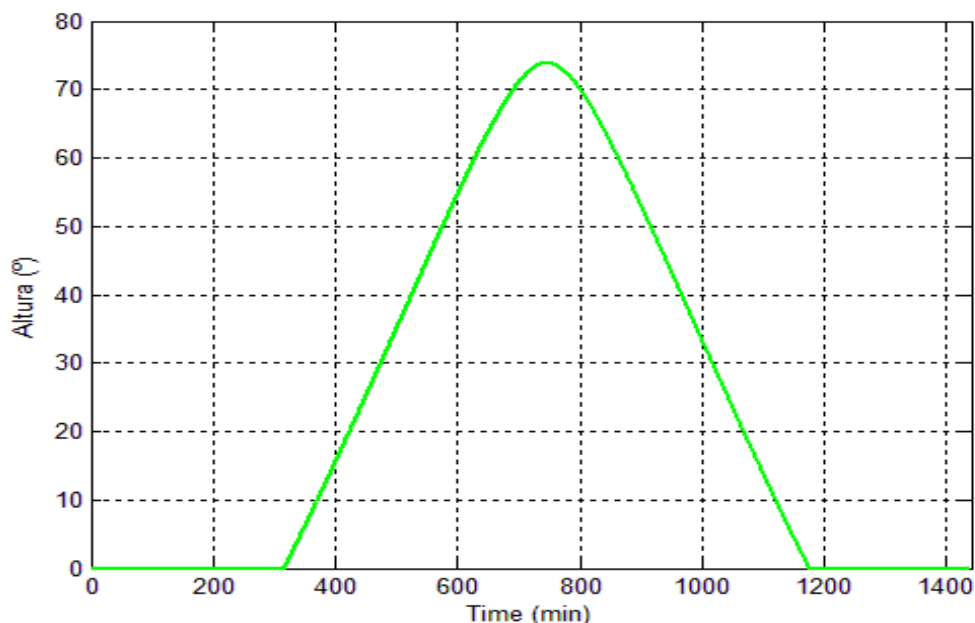


Figura 4. Altura solar a lo largo del día 187.

En la figura 4, vemos representada la evolución de la altura solar a lo largo de un día, hay que decir que los valores de la altura solar no coinciden a lo largo del año. Debido a los movimientos de los astros, los máximos valores de altura solar, varían de un día a otro.

Aun así, vemos que la altura solar es simétrica respecto a la mitad del día. Como la irradiancia directa está directamente relacionada con la altura solar, podemos ver la forma que tendrá la DNI. Al igual que vimos en la figura 1, la DNI es también aproximadamente simétrica, pero esto ocurre cuando el día es despejado, ya que, con nubes, esto no se cumple.

Respecto a la altura solar, este valor empieza a crecer desde 0 cuando el sol aparece en el horizonte al amanecer. Al medio día solar, la altura solar alcanza su valor máximo, volviendo a 0 en el ocaso.

Para el análisis de viabilidad de las Plantas Solares de Concentración (CSP), se suele simular una serie anual de valores de DNI característica del comportamiento a largo plazo del recurso en el emplazamiento y así tratar de conocer la producción de la planta anual más probable. Además de otros años que representan escenarios más desfavorables. Algunos modelos de simulación de plantas utilizados en proyectos comerciales, requieren que estas series estén al menos en resolución 10 minutal [3].

Para la generación de estas series características, es habitual utilizar datos de DNI estimados a partir de imágenes de satélite. Estos datos, suelen proporcionarse a escala horaria. Debido a esto, es necesario generar a partir de ellos una serie de valores con resolución diezminutal o incluso, si es posible, la generación de datos con resolución minutal.

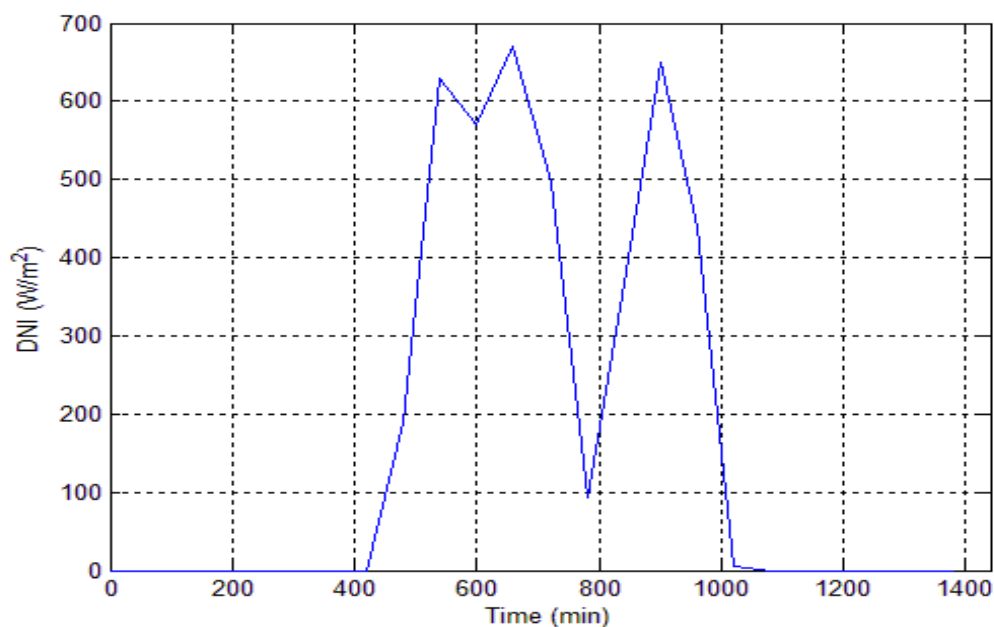


Figura 5. Valores DNI escala horaria.

En este caso, en la figura 5, mostramos los valores de DNI con resolución horaria para un día dado. A continuación, en la figura 6, mostramos los datos de DNI para el mismo día, pero con una resolución diezminutal.

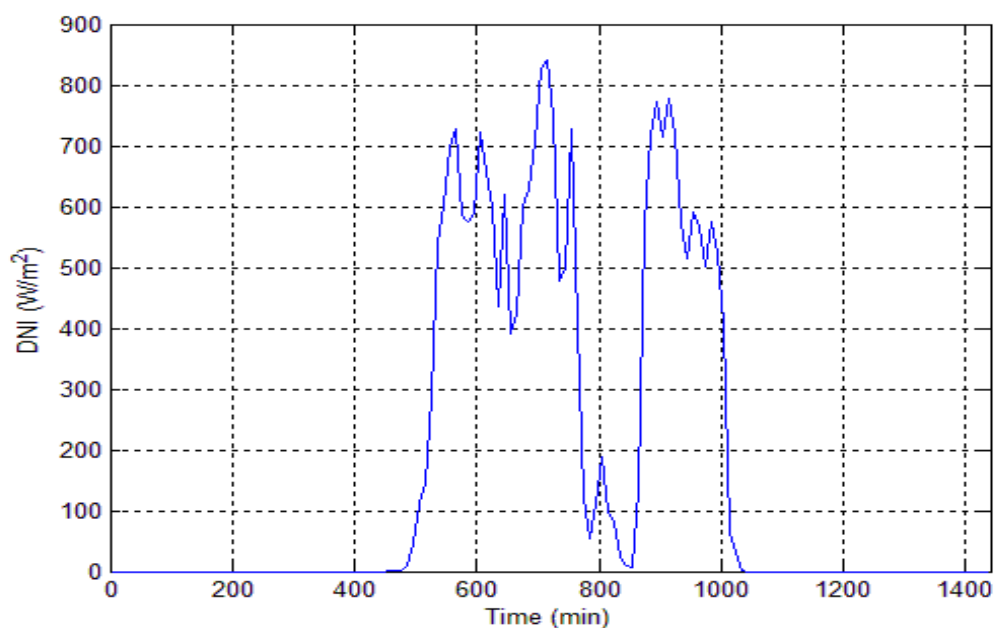


Figura 6. Valores DNI escala diez minutil.

Como hemos podido observar en la comparación entre la figura 5 y 6, el aumentar la resolución temporal, nos da mucha más información sobre los valores de DNI y su variabilidad a lo largo de un día, ya que la disminución de esta resolución suaviza los gradientes de DNI provocados por el paso de nubes. Es por ello, que algunos modelos de simulación utilizan series de valores de

DNI a una escala diezminutal, menor que la horaria. De igual modo, ampliar a escala minutal, da más información que la escala horaria o diezminutal, como se demuestra en la figura 7.

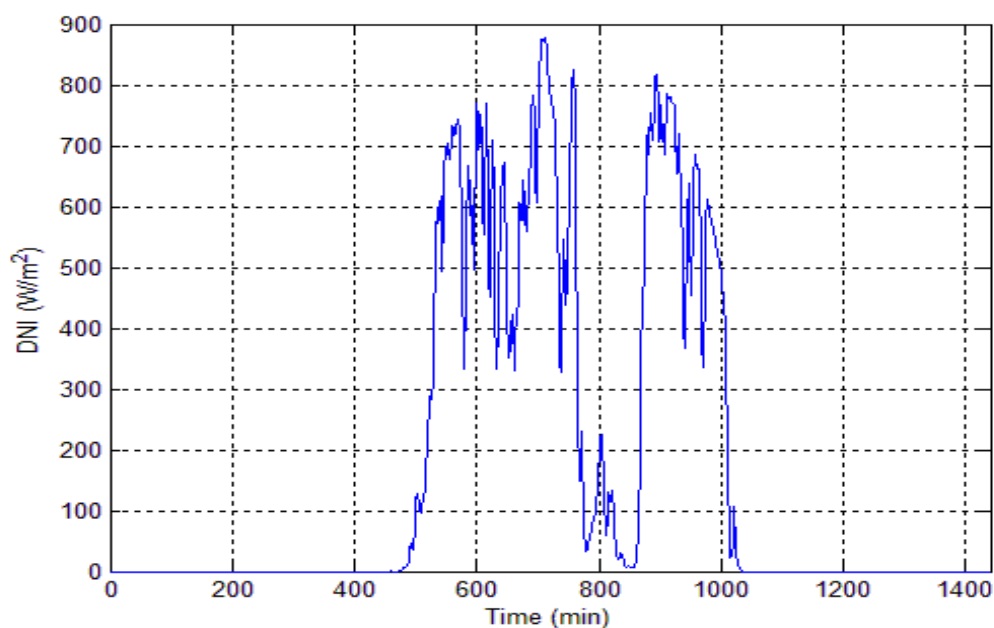


Figura 7. Valores DNI escala minutal.

Esto, lleva a la necesidad de desarrollar métodos que generen series de datos de alta resolución temporal a partir de datos horarios. Uno de estos métodos es el presentado por Fernández Peruchena [1], que consiste la generación de valores de DNI en escala 10 minutal a partir de valores horarios aplicando una novedosa técnica basada en el empaquetamiento de días medidos en escala 10 minutal. Este empaquetamiento consiste en la adimensionalización de la curva de DNI diaria, tanto en valores temporales como en valores de DNI.

Posteriormente, estos días empaquetados, son utilizados para ampliar la frecuencia temporal de días registrados en valores horarios, desempaquetando los días tanto en tiempo como en magnitud. La selección del día y el proceso de desempaquetamiento dependerá de los valores horarios de DNI de partida. En el apartado 4, se presenta una descripción más detallada de la metodología.

3 OBJETIVOS

El objetivo del presente trabajo es la aplicación y validación de la técnica de empaquetamiento de series diarias de datos de radiación directa normal a la base de datos del GTER.

Esta validación se llevará a cabo a partir del análisis de varios índices estadísticos. A su vez, este método de generación de series de mayor resolución temporal, será comparado con el método presentado por Larrañeta [3].

4 MÉTODO DE EMPAQUETAMIENTO DE DÍAS

Para obtener datos de DNI de alta frecuencia, C.M. Fernández Peruchena [1] ha desarrollado un método que consiste en que, a partir de datos horarios, es posible obtener datos de DNI con resolución minutal.

El proceso de obtención de estas series de datos se centra en el uso de curvas de DNI medidas, de alta frecuencia, y que han sido adimensionalizadas. Estas, serán la base de la generación de meses o años para un lugar dado. Es por ello que para aplicar este método es necesario contar con una base de datos previa a escala minutal o diezminutal.

La serie de datos de alta resolución temporal de la que se parte, no está adimensionalizada. Es por ello que para aplicar este método hay que seguir una serie de pasos previos, desde la obtención de las curvas adimensionalizadas, hasta la generación de un año sintético. Estos pasos se pueden agrupar en 3 puntos:

1. Cálculo de la envolvente de DNI de cielo despejado. Esto se consigue mediante el ajuste de dos parámetros (E_0, β) pertenecientes al modelo ASHRAE, obteniendo una curva lo más próxima posible a los valores máximos de un día en el que no existen nubes. El modelo exponencial ASHRAE se define mediante la siguiente expresión:

$$DNI = E_0 \cdot \exp\left(\frac{-\beta}{\sin(\alpha)}\right)$$

Donde α es la elevación solar, E_0 es la irradiancia aparente extraterrestre y β es el parámetro de extinción.

2. Cálculo de las curvas adimensionales de DNI. Las curvas adimensionales de alta frecuencia consistirán en la adimensionalización tanto de los valores temporales, como de los valores de DNI.

Para ello hay que conocer la duración del día, es decir el total de tiempo que hay irradiancia, y los valores máximos de DNI, valor de cielo despejado, para cada uno de esos instantes. Una vez conocido esto, el valor de cualquier instante se dividirá por la duración total de ese día, y el valor de DNI para ese instante se dividirá por el valor de cielo despejado para ese mismo momento.

Con esto se obtienen curvas cuyos ejes se mueven entre 0 y 1. Siendo en el caso de DNI, 1 un valor medido igual al valor de cielo despejado y 0 un valor nulo respecto al cielo despejado. Y en el caso del tiempo, 0 corresponde a la salida del sol y 1 corresponde al ocaso. A este proceso se le llamará empaquetado.

3. Generación de series de DNI de alta frecuencia. Una vez aplicado los dos pasos anteriores, se obtiene un año completo adimensionalizado en el que todos los días tienen la misma escala vertical y horizontal.

El método propone que, a partir de datos horarios, un algoritmo encuentre el día que más se parezca en términos de DNI de entre todos los días empaquetados.

El algoritmo parte con los datos de DNI correspondientes a un día de resolución temporal horaria. Como introducimos, estos días de los que parte, no tienen la misma duración en lo que a tiempo se refiere, es decir, un día puede tener más horas de sol que otro. Por ello, hay que conocer la duración de ese día en escala horaria, para poder comparar con los días empaquetados. Una vez conocido, dividir el eje horizontal, correspondiente al temporal, entre la duración del día del que partimos. Con ello conseguimos que la duración del día empaquetado a comparar sea independiente de la duración del día de cuando fue empaquetado, lo que nos permite comparar con la base de datos de los 365 días.

Es por ello, que podemos decir que cuando partimos de un día, tenemos 365 combinaciones posibles. De igual modo que hemos hecho ajustando la duración del día, también hay que ajustar el eje vertical, correspondiente a los valores de DNI.

Para ello hay que tener previamente una envolvente de cielo despejado con la que desempaquetar. El método de Fernández Peruchena [1], propone la utilización de los valores de E_0 y β que se han calculado antes de empaquetar, y utilizar de manera independientemente E_0 de β . Por ello, si antes contábamos con 365 pares de E_0 y β , ahora contamos con 365 E_0 para cada valor de β y viceversa. Esto nos lleva a pensar que las combinaciones numerosas y esto supondría un coste computacional muy elevado

Para obtener la combinación más adecuada de curva adimensional y de datos empaquetados, el algoritmo buscará el día más adecuado en términos de distancia euclídea, comprobando en cada caso esta distancia con todas las combinaciones de E_0 y β .

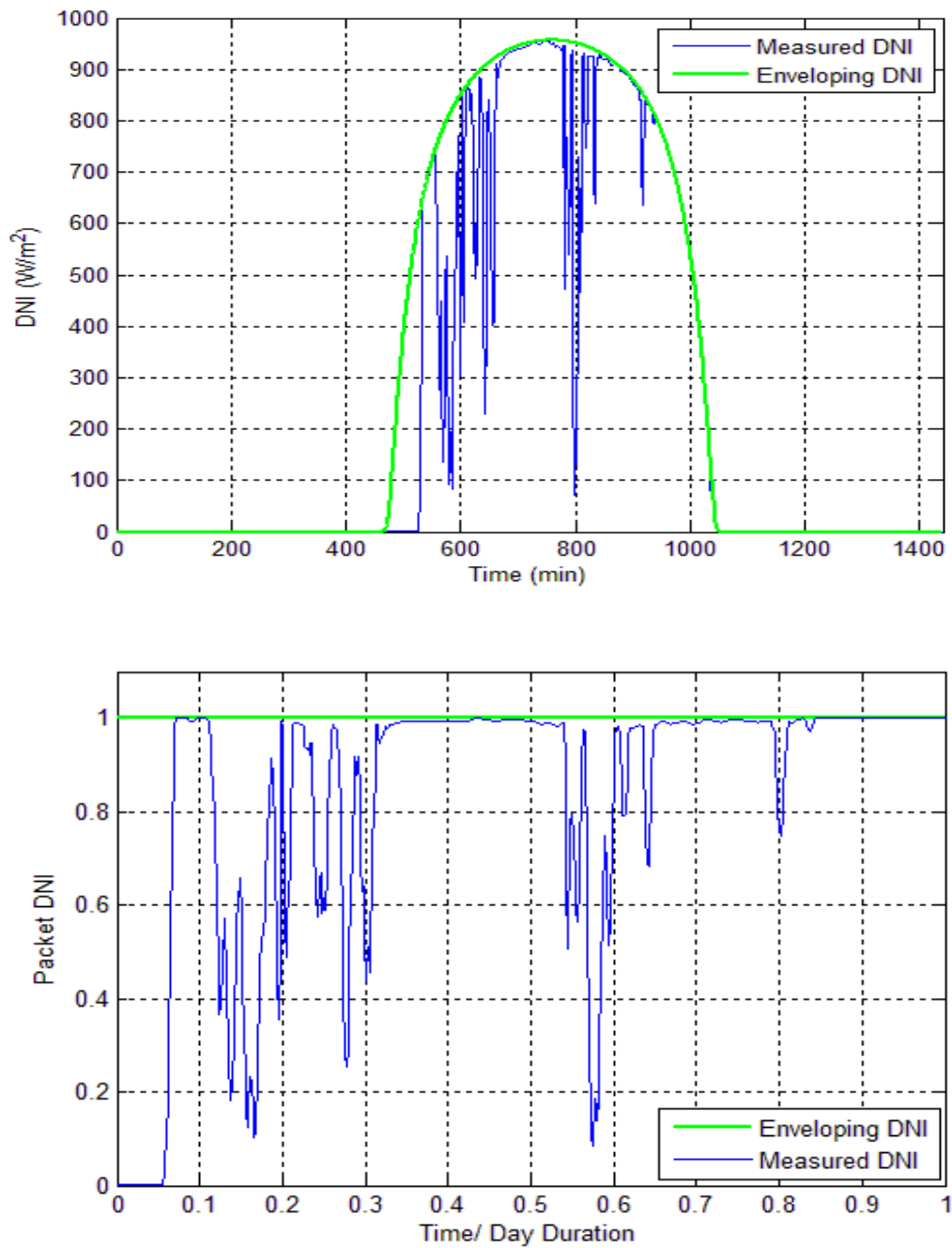


Figura 8. Curva datos DNI minutales con curva de día despejado y día adimensionalizado.

En la figura 8 podemos ver, en la parte superior, los valores medidos de DNI en azul y su correspondiente envolvente en verde, y en la parte inferior, ese mismo día empaquetado.

5 APLICACIÓN DEL MÉTODO

Para corroborar que el método ha sido implementado correctamente, inicialmente, se ha aplicado a la misma base de datos utilizada por Fernández Peruchena. La base de datos pertenece a Carpentras Baseline Surface Radiation Network (BSRN) del mes de agosto de 2010.

5.1 CÁLCULO DE LA ENVOLVENTE DE CIELO DESPEJADO

Siguiendo el primer paso del método, se calculan los valores de E_0 y β con el modelo exponencial ASHRAE, definido por la siguiente expresión:

$$DNI = E_0 \cdot \exp\left(\frac{-\beta}{\sin(\alpha)}\right)$$

Para ello, se parte de los datos minutales correspondientes a un día del mes de agosto de 2010 de la Carpentras Baseline Surface Radiation Network. En este caso, y a modo de ejemplo, tomaremos el día 22 de agosto:

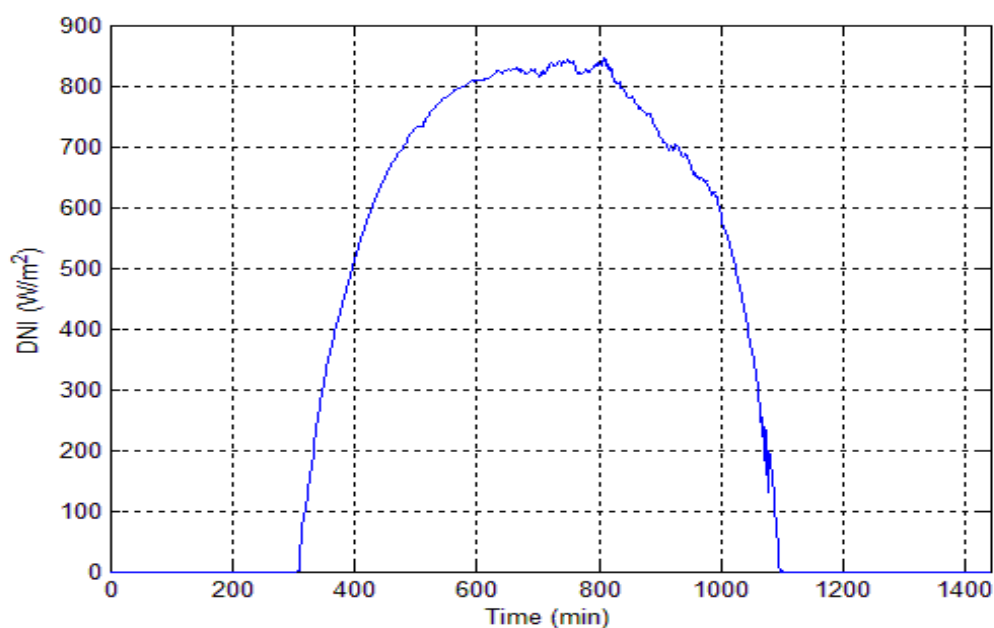


Figura 9. Curva datos minutales DNI, 22 de agosto 2010.

Como vemos en la figura 9, al partir de datos minutales contamos con un total de 1440 puntos con sus respectivos valores de DNI. Para la obtención de los valores de E_0 y β , debemos sólo tener en cuenta los valores de DNI que representen al día despejado, por ello, hay que aplicar un filtro con el objetivo de obtener esos valores.

Se aplican un total de 3 filtros diferentes:

1. Valores con baja elevación solar: El primer filtro consiste en eliminar aquellos valores que tengan una baja elevación solar. Para ello, eliminaremos aquellos valores a los que correspondan un valor de elevación solar menor de 20° . Estos valores son filtrados debido a fenómenos de refracción y a que para elevaciones bajas es cuando la atenuación atmosférica es mayor, ya que tiene mayor recorrido hasta el suelo. La utilización de estos valores podría introducir ruidos en el resultado, y por ello se eliminan.
2. Puntos con bajo valor de DNI: El segundo filtro consiste en la eliminación de los valores que tengan un valor medido de DNI menor de 300 W/m^2 .
3. Eliminación de puntos que contengan un valor menor de DNI respecto a otro con menor elevación solar y un valor de DNI mayor. El valor de DNI está directamente relacionado con la elevación solar, a mayor elevación solar, mayor es el valor de DNI.

Supongamos que para el día 22 de agosto contamos con dos puntos distintos:

- Punto 1: Corresponde a un valor de DNI de 750 W/m^2 , y una elevación solar de 40° .
- Punto 2: Corresponde a un valor de DNI de 800 W/m^2 , y una elevación solar de 35° .

Como podemos observar la elevación solar del punto 1 es mayor que la del punto 2, sin embargo, el valor de DNI es menor en el punto 1 que en el punto 2.

En un día totalmente despejado. El punto 1 al corresponderle una mayor elevación solar de 40° , tendría que tener un valor de DNI mayor al que le corresponde al punto 2, que tiene una elevación solar de 35° . Esto es debido a que en el momento en el que se tomaron los datos del punto 1 pasó una nube, por lo que este punto no sería útil a la hora de calcular la envolvente de cielo despejado.

Una vez aplicados los 3 filtros a los datos del día 22 de agosto, se obtienen una serie de puntos que corresponderían al cielo despejado en ese día, como se muestra en la figura 10:

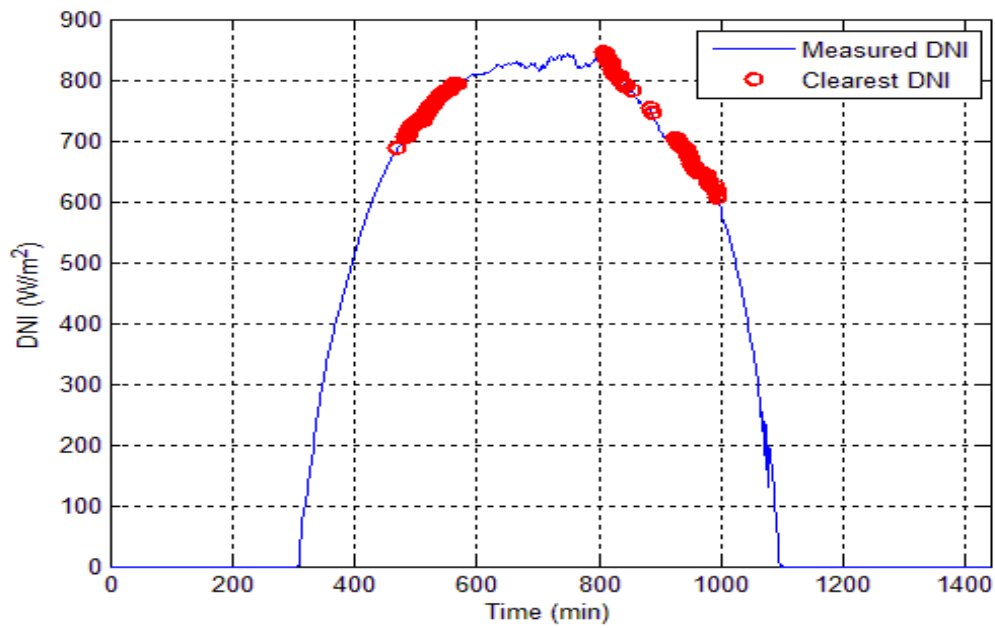


Figura 10. Curva datos minutales DNI con puntos de cielo despejado, 22 de agosto 2010.

Una vez calculados los puntos de cielo despejado, para obtener los valores de E_0 y β calcularemos la aproximación lineal. Para ello, tomamos la ecuación del modelo exponencial ASHRAE:

$$DNI = E_0 \cdot \exp\left(\frac{-\beta}{\sin(\alpha)}\right)$$

Y despejamos el valor β aplicando logaritmo neperiano:

$$\ln(DNI) = \ln(E_0) - \beta \cdot \left(\frac{1}{\sin(\alpha)}\right)$$

Aplicando esta ecuación a los puntos de cielo despejado obtenidos mediante los filtros anteriores se obtiene:

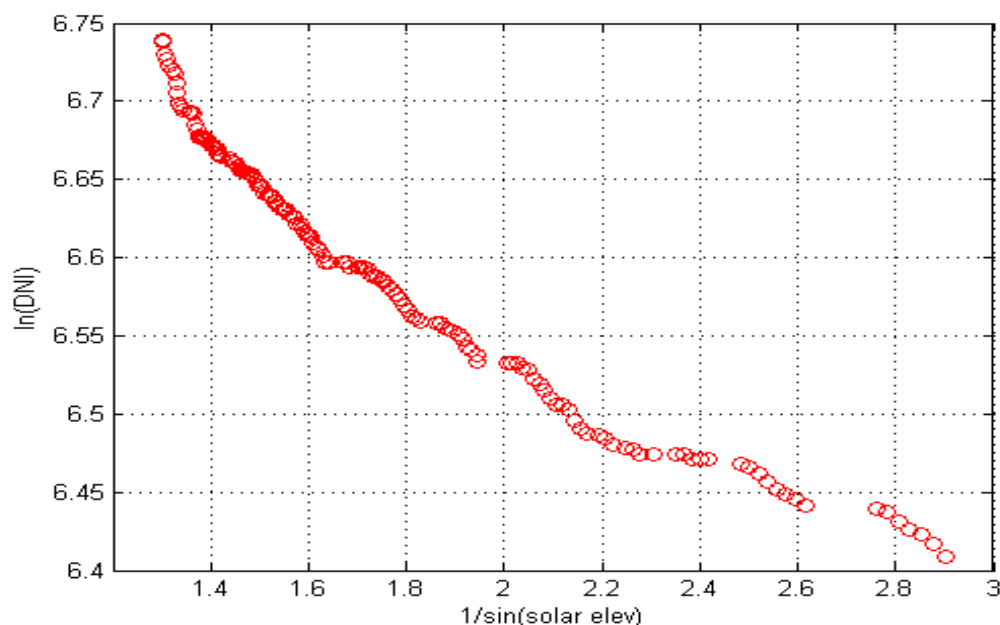


Figura 11. Curva puntos cielo despejado.

La aproximación lineal se trata de un ajuste robusto. El ajuste robusto es una forma de análisis de la regresión, que a la hora de calcular la aproximación lineal no tiene en cuenta los valores atípicos de la serie a ajustar. Esto quiere decir que elimina los elementos que no siguen el patrón de las demás observaciones. El entorno de trabajo que estamos utilizando es Matlab, el cual cuenta con varios métodos de ajuste lineal robusto: Andrews, Bisquare, Cauchy, Fair, Huber, Logistic, Ols, Talwar y Welsch.

Como estamos reproduciendo el método de Fernández Peruchena, el cual se basa en el entorno R, tenemos que analizar qué método de aproximación de Matlab se acerca más a los valores obtenidos para la base de datos de Carpentras para agosto de 2010. Como conocemos los resultados, ya que nos han sido proporcionados, aplicaremos todos los ajustes lineales hasta hallar el que más se ajusta a los resultados de los que partimos, que se connotarán como 'MFP'.

	MFP	Andrews	Bisquare	Cauchy	Fair	Huber	Logistic	Ols	Talwar	Welsch
β	0.204	0.234	0.234	0.206	0.202	0.204	0.203	0.197	0.195	0.233
E%	-	14.7	14.7	0.9	0.9	0	0.5	3.4	4.4	14.2
E_0	1101.9	1201.05	1200.8	1107.6	1099.5	1102.8	1102.7	1091.9	1088.9	1198.7
E%	-	8.9	8.9	0.5	0.2	0.1	0.1	0.9	1.2	8.8

Tabla 1. Valores E_0 y β día 22 de agosto.

A la vista de los resultados, tanto el método Huber como el método Logistic, se acercan bastante al original, en este caso optaremos por el método Huber.

Una vez hecho el ajuste lineal, como hemos visto se obtiene un valor de E_0 y β , pero este es un ajuste de los puntos:

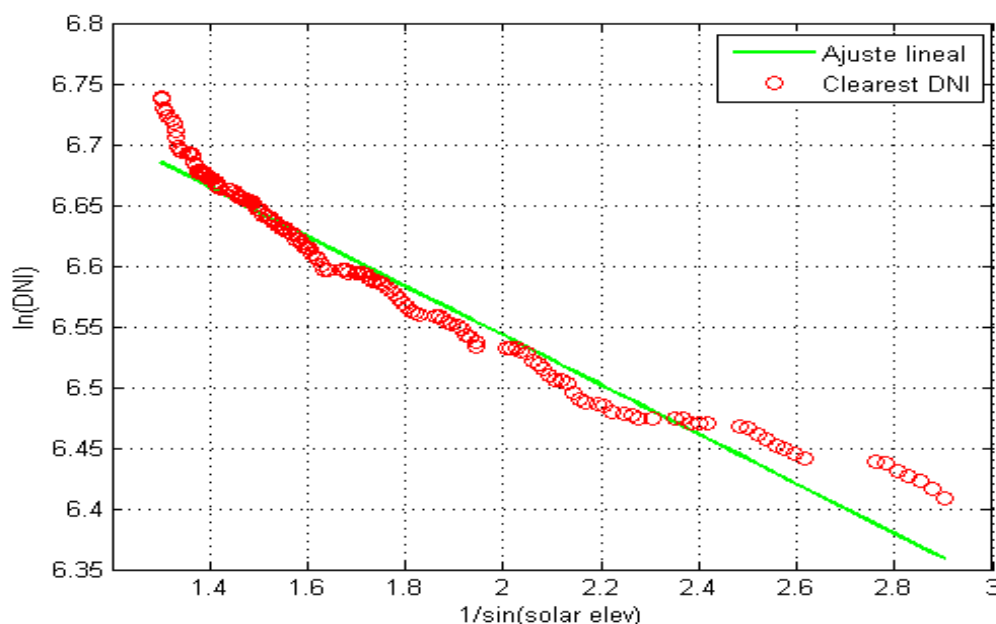


Figura 12. Ajuste lineal robusto.

Para que los valores de E_0 y β correspondan a la envolvente máxima, esta se debe adecuar al punto superior más lejano en términos de distancia. Para el cálculo de esta distancia, trazamos una recta perpendicular a la recta del ajuste lineal que pase por el punto de cielo despejado. El valor de esta distancia será la longitud que existe entre la línea de ajuste lineal y el punto de cielo despejado pasando por esa recta perpendicular al ajuste. La distancia será positiva si el punto se encuentra por encima de la línea de ajuste robusto y negativa en caso contrario.

Esta distancia se calcula para cada punto que representa un valor de cielo despejado, y se busca el valor positivo máximo, que corresponderá a la máxima distancia. El objetivo de esto es hacer pasar a la recta de ajuste lineal por ese punto, lo que haría que la envolvente para ese día sea la máxima.

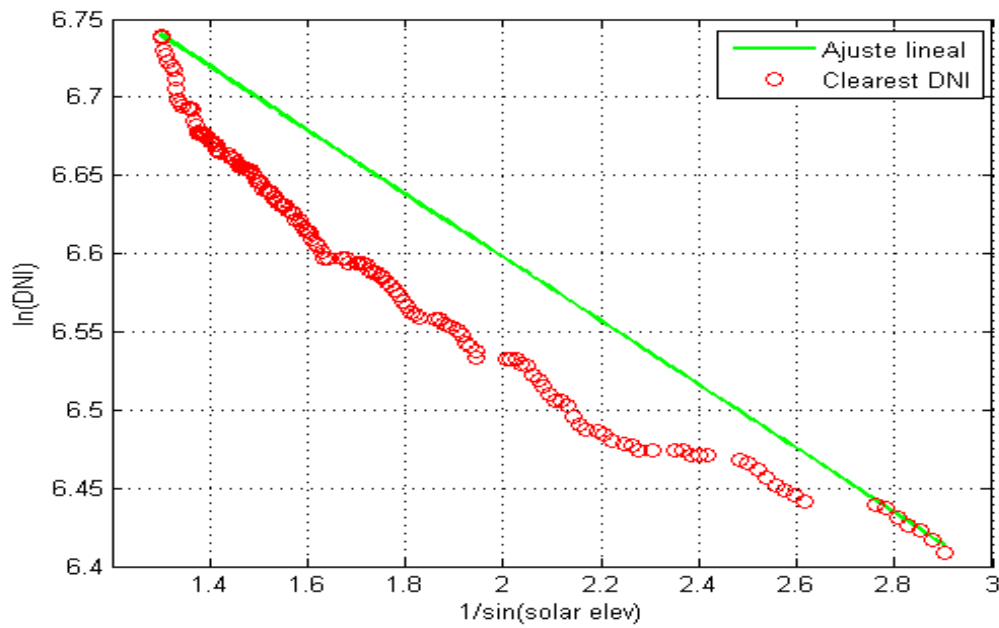


Figura 13. Ajuste lineal robusto de cielo despejado.

Una vez hecho el ajuste se obtiene que, para el día 22 de agosto de 2010, el valor de $E_0=1102.8$ y el valor de $\beta=0.204$.

Aplicamos estos valores a la fórmula del modelo exponencial de ASHRAE obteniendo así la envolvente de cielo despejado:

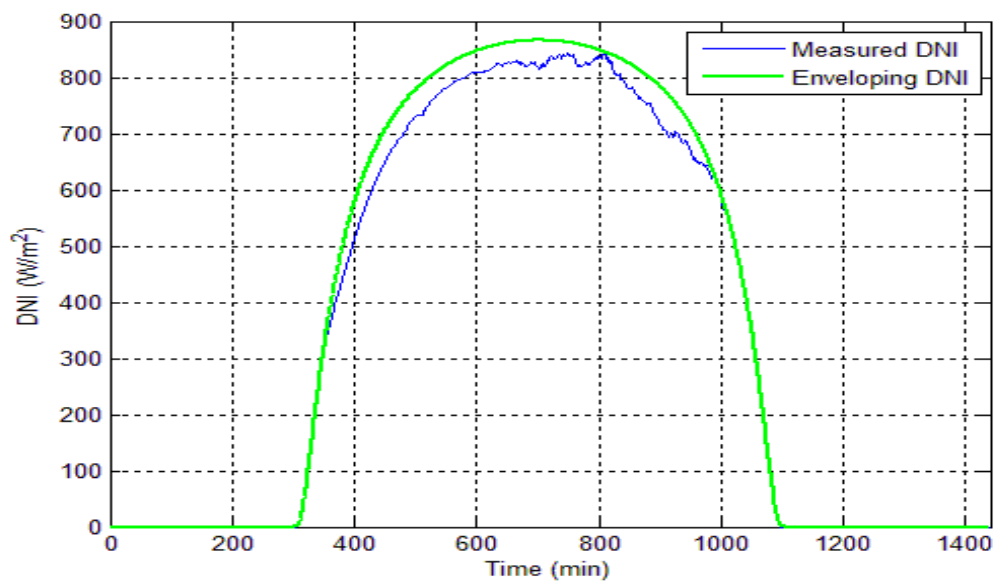


Figura 14. Datos minutales DNI y envolvente de cielo despejado.

El siguiente paso consiste en el empaquetado del día, obteniendo como ejes horizontales y verticales valores comprendidos entre 0 y 1.

El eje horizontal será el eje temporal que comprenderá entre 0 y 1, siendo 0 la salida del sol y 1 el ocaso. En el momento del empaquetado, supondremos que el sol saldrá cuando el valor de la elevación solar sea mayor de 6° y se pondrá cuando el valor sea menor de 6° . Teniendo en cuenta esto, el empaquetado final será el siguiente:

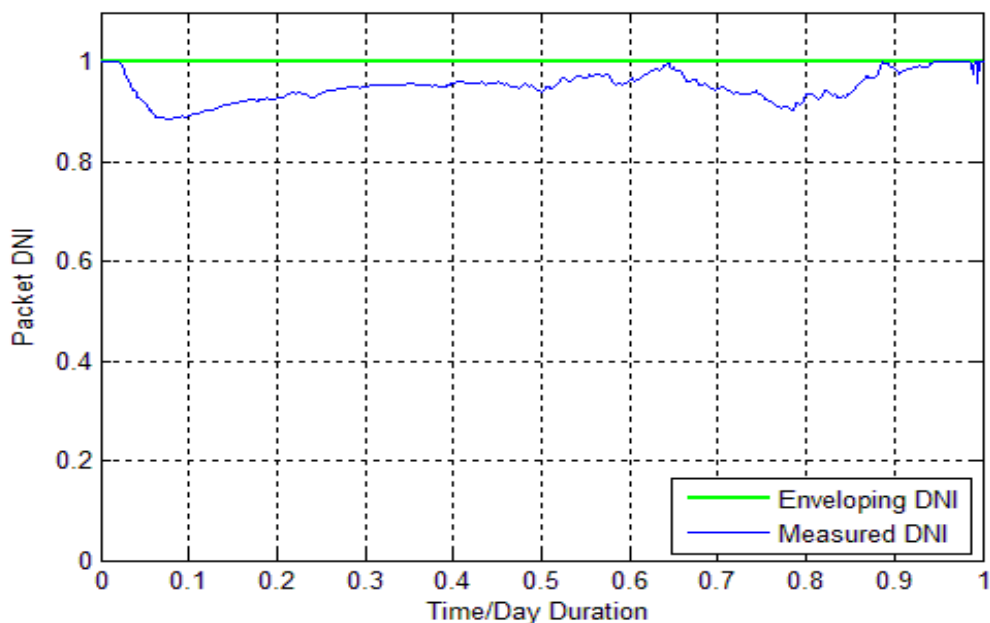


Figura 15. Día 22 de agosto adimensionalizado.

Como hemos podido comprobar hasta ahora, el empaquetamiento es independiente para cada día. Es decir, los días no duran lo mismo ni tienen los mismos valores máximos de DNI.

A pesar de ello, estos días una vez empaquetados, tienen que poder ser desempaquetados en una escala temporal distinta a la que fue empaquetada.

Un día puede, entre la salida del sol y el ocaso, tener 500 minutos de sol. Mientras que otro día puede tener solamente 300 minutos. En este caso, se podría empaquetar cada día con su duración en minutos, obteniéndose 500 y 300 valores respectivamente, pero esto añadiría una grave problemática a la hora de comparar estos días.

Por ello, procedemos a interpolar estos puntos para llevarlos a un total de 1000, es decir, todos los días acabarán en paquetes cuyos ejes horizontales y verticales van de 0 a 1 y el eje horizontal estará definido por 1000 puntos.

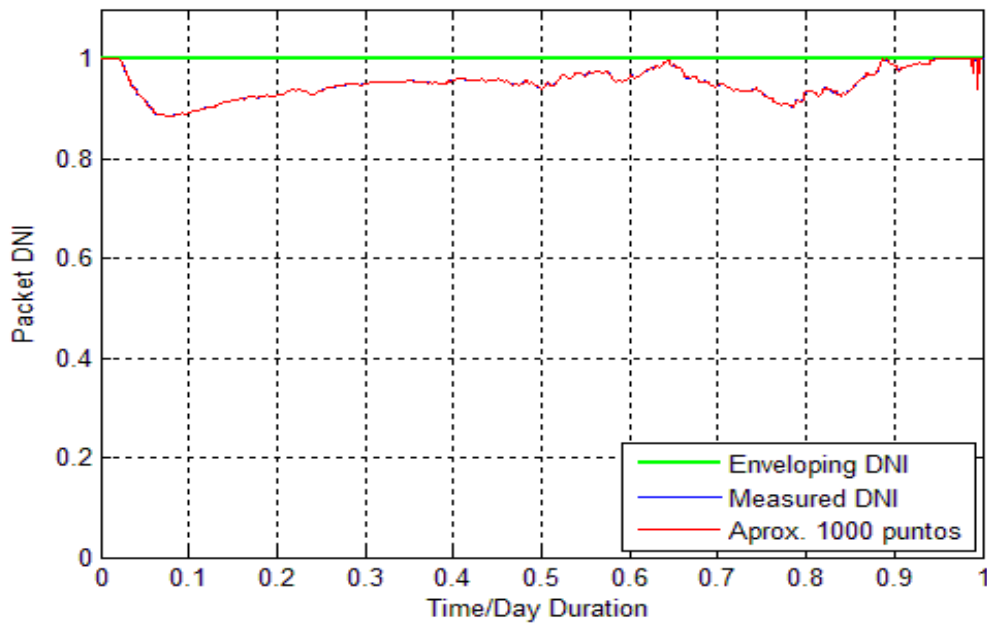


Figura 16. Día 22 de agosto adimensionalizado y aproximación de 1000 puntos.

Como vemos en la figura 16, la curva en rojo representa la aproximación de 1000 puntos, como observamos se ajusta a la curva dada, por lo que estos puntos son suficientes para aproximar el empaquetado.

5.2 ANÁLISIS DE SIMILITUD

Una vez aplicado el método, existían dos métodos distintos de ajuste lineal robusto, métodos Logistic y Huber. Para la elección del método adecuado, aplicamos ambos métodos a todo el mes de agosto y lo comparamos con los datos de partida:

	Andrews	Bisquare	Cauchy	Fair	Huber	Logistic	Ols	Talwar	Welsch
Error(%)	17.4	17.3	14.9	17.9	0.35	1.95	158.5	33.8	17.4

Tabla 2. Error E_0 y 6 mes de agosto.

En la tabla 2, se muestra el mayor error cometido, para el mes de agosto, respecto al modelo presentado por Fernández Peruchena [1] en término porcentual. Como observamos el método Huber es el que menos error comete, tan solo un 0.35%, por lo que finalmente este será el método de ajuste lineal elegido.

6 APLICACIÓN A LA BASE DE DATOS GTER

El objetivo de este proyecto, es la aplicación de este método a la base radiométrica del GTER, que mide los datos en Sevilla.

Para ello, hemos contado con la base de datos para el año 2014 en resolución horaria, a partir de la cual, vamos a generar el año sintético. Aunque también contamos con las mediciones de DNI minutas y diez minutas para validar la metodología.

Para generar el año a partir de los datos horarios del 2014, procederemos a empaquetar la base de datos del 2000 al 2012 en base diezminutal. Esto nos dará una base de datos de 13 años empaquetados.

Para visualizar el proceso de empaquetado de los 13 años con los que contamos, vamos a aplicar el método para un día y en este caso, a modo de ejemplo, en valores minutas aunque todo el proceso llevado a cabo en este proyecto se hará con datos diezminutales. Aplicando el método, obtendremos valores de E_0 y β para un solo día:

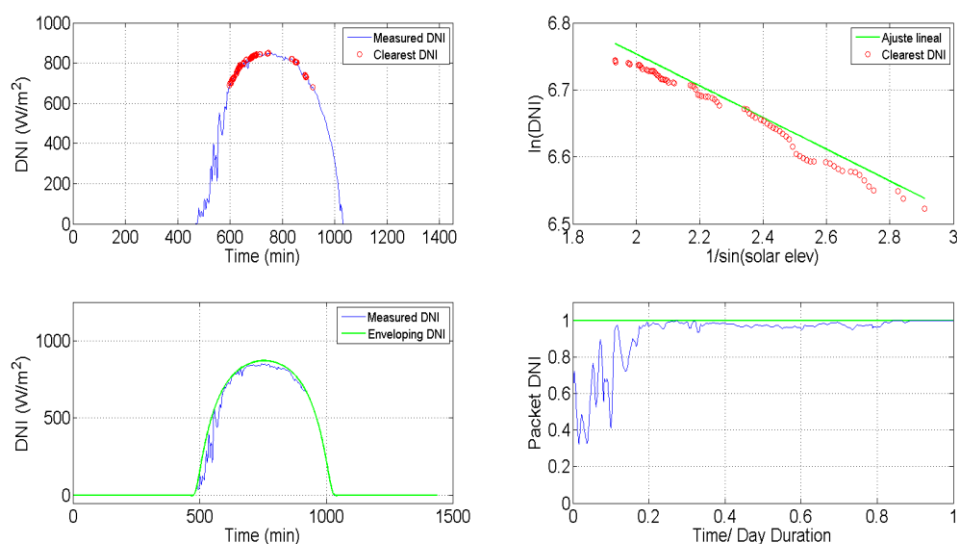


Figura 17. Aplicación método empaquetamiento. 7 de enero.

Al final del proceso de empaquetamiento obtendremos cerca de 9500 valores de E_0 y 9500 valores de β , junto a ello, un vector de 4.749.000 puntos, debido a los 1000 puntos por día en los que hemos empaquetado.

6.1 PROBLEMÁTICA DEL MÉTODO

Al aplicar el método, puede ocurrir lo que observamos en las siguientes figuras:

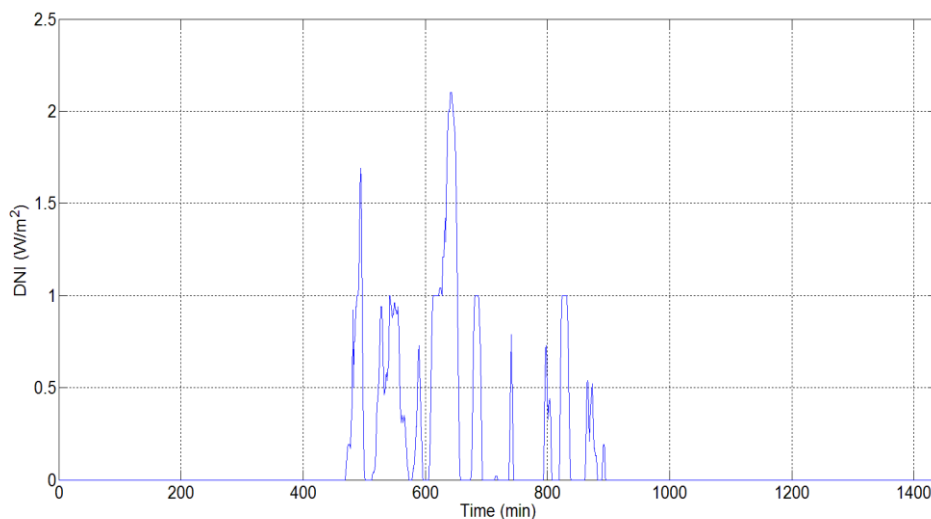


Figura 18. Aplicación método empaquetamiento. 2 de enero.

Como podemos observar, en la figura 18, el día 2 de enero se trata de un día nublado. En él, los valores de DNI no superan los 2.5 W/m^2 . Por lo que, a la hora de calcular la envolvente, no existirían puntos con los que llevar a cabo el ajuste lineal. Debido a que no superan los 3 filtros que permiten escoger los datos de DNI de cielo despejado.

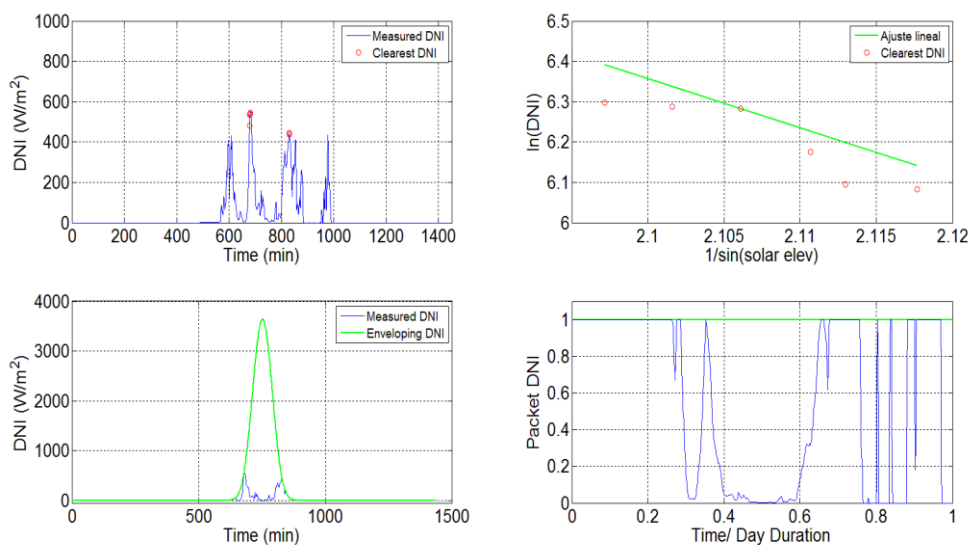


Figura 19. Aplicación método empaquetamiento. 4 de enero.

Como observamos de nuevo en la figura 19, el día 4 de enero se trata de otro día nublado, en el que sí existen puntos que permiten hacer el ajuste lineal, pero estos puntos no permiten que el ajuste lineal robusto sea correcto, ya que la curva de cielo despejado llega a valores superiores a 3000 W/m^2 .

Por lo que llegamos a la conclusión de que el ajuste lineal robusto necesita un mínimo de puntos para que se ajuste a la curva de DNI medida y demuestra que el método no da la posibilidad de empaquetar los días nublados, por lo que tampoco permitiría desempaquetarlos, imposibilitando la generación de un año completo.

6.2 APLICACIÓN DEL MÉTODO A DATOS HORARIOS

El método de Fernández Peruchena [1] consistía en la generación de datos de DNI de alta resolución a partir de datos horarios. Por ello, para la generación del año partimos de datos horarios de 2014.

Los días empaquetados tienen ejes cuyos valores varían entre 0 y 1. Por lo que, para desempaquetar los valores del eje vertical, hay que contar con el valor máximo de DNI para cada instante. Esto se consigue mediante la utilización de la envolvente de cielo despejado.

Por ello, hemos optado por aplicar el método de cálculo de envolvente a los datos horarios, y obtener nuevos valores de E_0 y β a partir de esos datos horarios. A estos datos hay que aplicarle de nuevo los 3 filtros del método de Fernández Peruchena. Pero como comprobaremos, al ser una escala temporal horaria, los datos de los que partimos cuentan con muy pocos puntos, por lo que, en este caso, el segundo filtro desechará los valores de DNI menores a 300 W/m^2 .

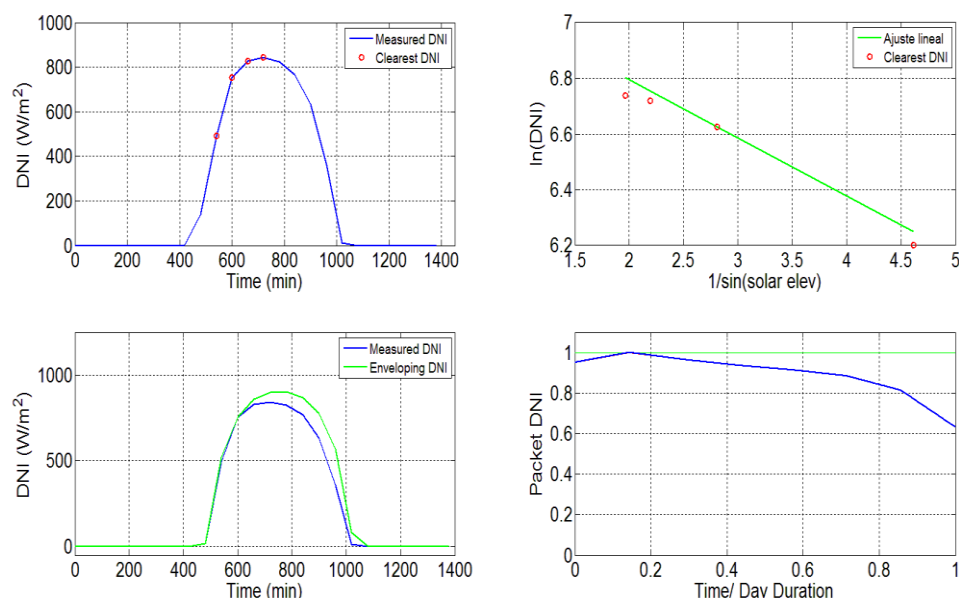


Figura 20. Aplicación método empaquetamiento datos horarios. 7 de enero.

Como vemos en la figura 20, a pesar de contar con sólo 4 puntos, este método consigue calcular una envolvente de cielo despejado que se ajusta a la curva de datos horarios.

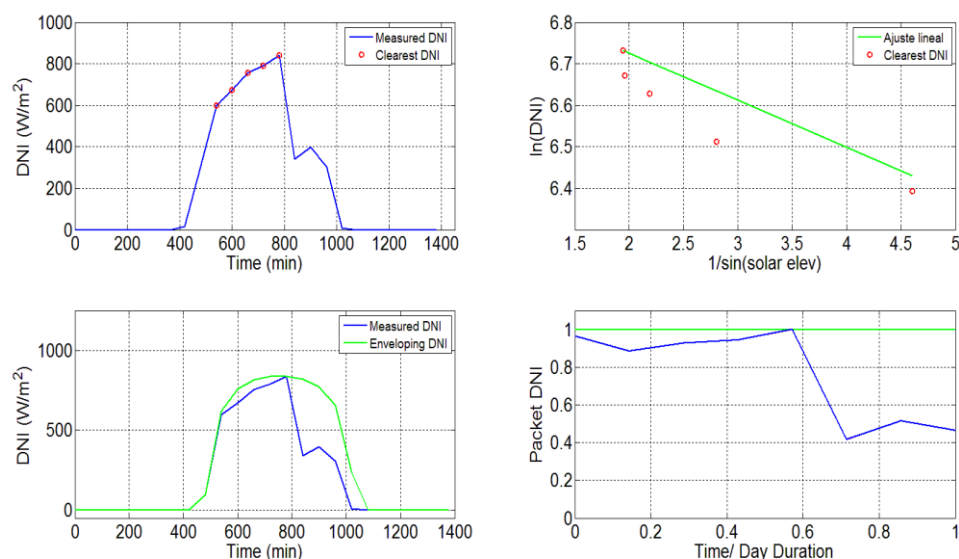


Figura 21. Aplicación método empaquetamiento datos horarios. 8 de enero.

En este caso en la figura 21, observamos un día en el que ha habido cierta nubosidad y, aun así, la envolvente de cielo despejado se ajusta a los datos horarios de partida.

A pesar de ello, nos encontramos con el mismo problema que en caso de datos minutales:

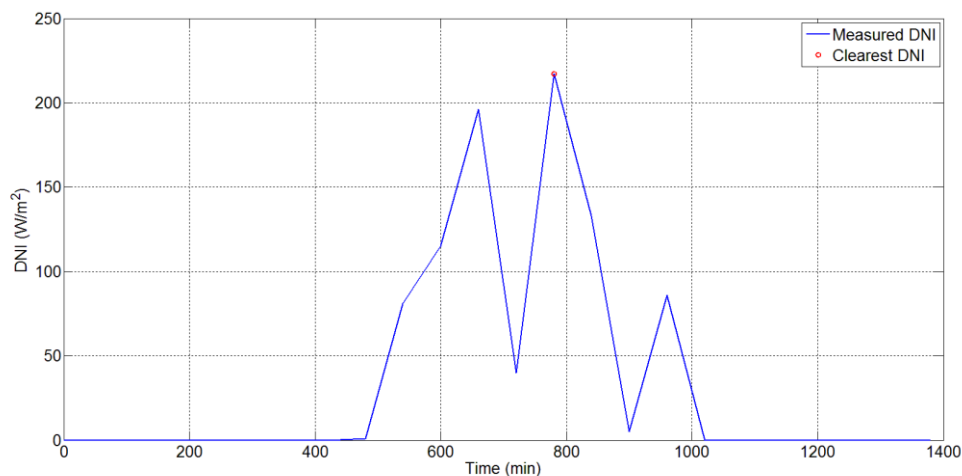


Figura 22. Aplicación método empaquetamiento datos horarios. 4 de enero.

La figura 22 representa que hay días en los que no se alcanzan valores mínimos de DNI como para calcular envolvente de cielo despejado. Por lo que resulta imposible comparar y buscar en los días empaquetados aquel que más se ajuste.

En este caso sólo es posible obtener 1 punto de cielo despejado, debido a que, obteniendo datos horarios, no se consiguen tantos puntos como en el caso de los datos minutales.

Otro problema que presenta el aplicar este método a los valores horarios es el siguiente:

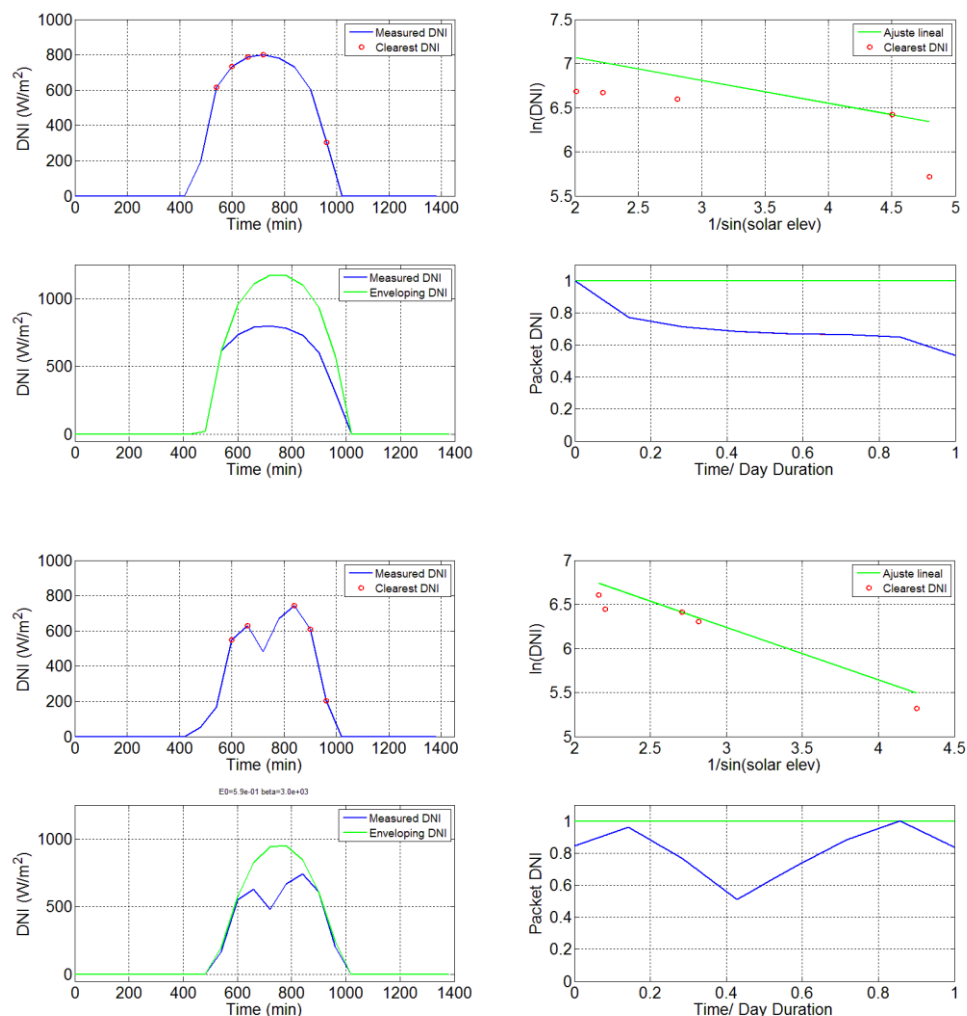


Figura 23. Aplicación método empaquetamiento datos horarios. 22 de diciembre (superior). 5 de enero(inferior).

Como vemos en la figura 23, existen días en los que el ajuste lineal robusto no consigue obtener una curva de cielo despejado que se acerque a los valores reales de los datos horarios. Para solucionar este último problema hemos optado por obtener puntos a escala minutal a los que aplicar este método.

6.2.1 Ampliación minutal de datos horarios

Este método tiene como objetivo, conseguir que existan más puntos para obtener un correcto ajuste de la curva de cielo despejado con los datos horarios. Consiste en simplemente interpolar linealmente los valores horarios, con el objetivo de pasar de 24 puntos que analizar, a 1440 puntos.

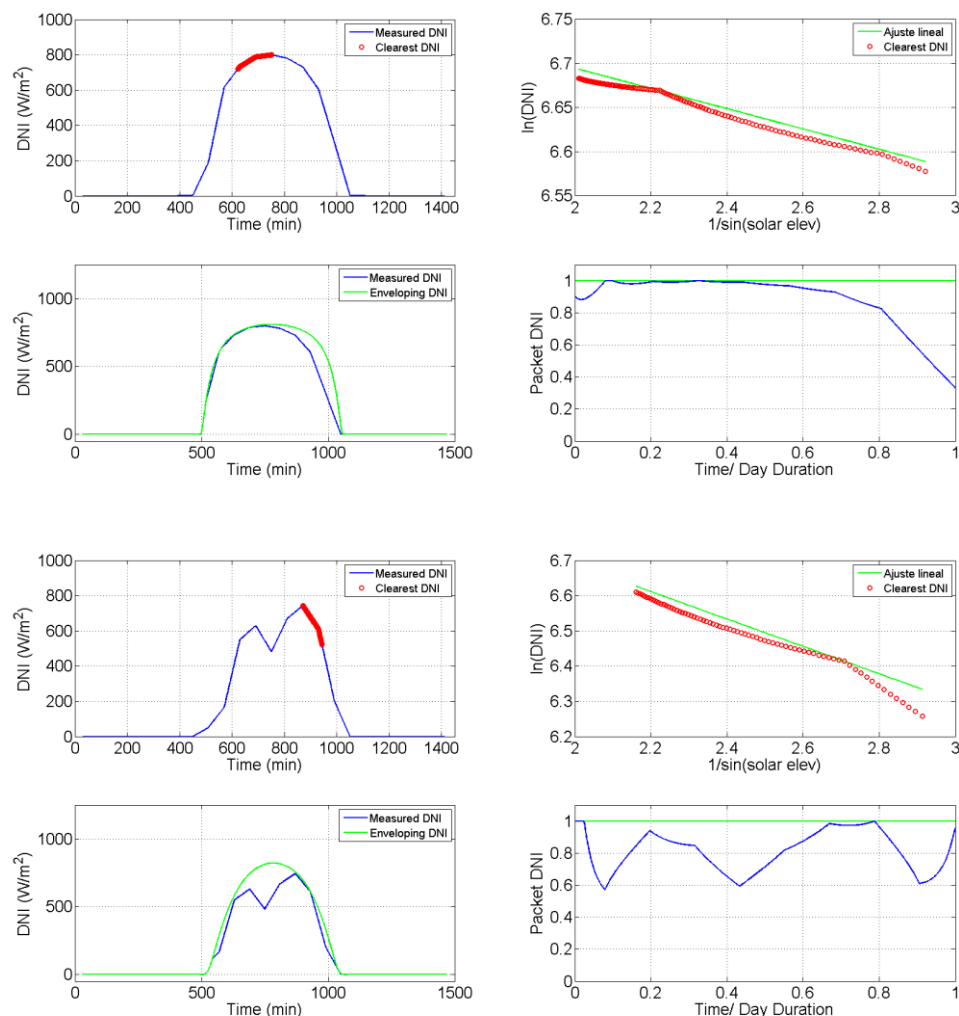


Figura 24. Aplicación método empaquetamiento datos horarios con ampliación minutal. 22 de diciembre (superior). 5 de enero(inferior).

Una vez aplicado el método podemos observar en la figura 24, cómo las curvas de envolvente ahora se ajustan más a los datos horarios, con lo que, a la hora de buscar en los valores adimensionales, se ajustará el desempaquetado más a los valores reales minutales.

Aplicamos este método al día 8 de enero, el cual calculamos antes para demostrar que ampliar los datos a escala minutal no afecta a los días con los que ya se calculaba la envolvente de manera ajustada:

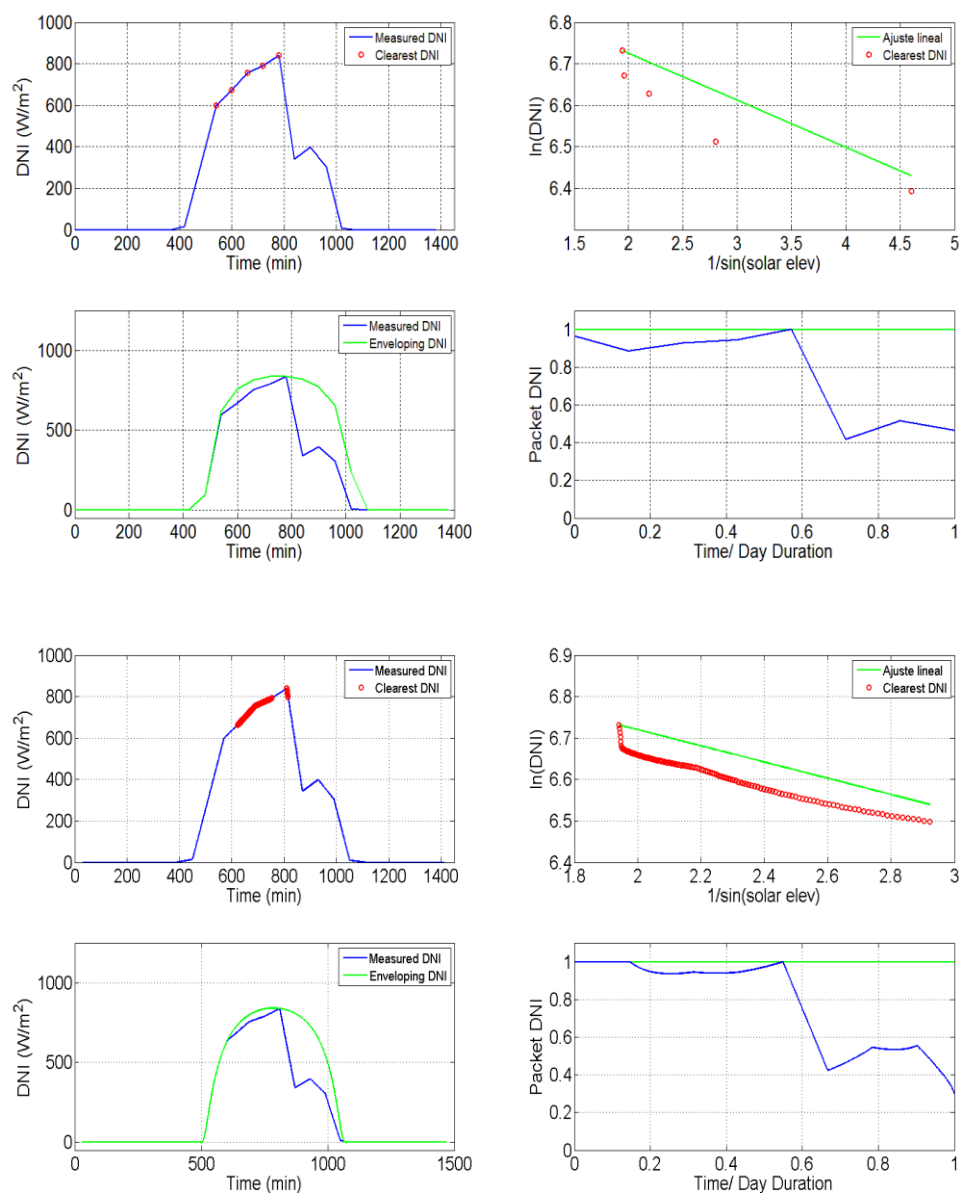


Figura 25. Cálculo de envoltorio día 8 de enero. Aplicación sin ampliación horaria (Arriba). Aplicación con ampliación horaria (Abajo).

Como observamos en la figura 25, el implementar la ampliación minutal de los datos horarios, no afecta a los días para los que la envoltorio sí se ajustaba correctamente a esos valores, por lo que aplicar la mejora a este método, no supondría una desventaja respecto al método original partiendo de los datos horarios.

6.3 EMPAQUETAMIENTO DE DÍAS NUBLADOS

Como hemos observado tanto en el proceso de cálculo de envolvente de los datos horarios como en los datos minutales, existe la problemática de que de los días nublados no se pueden obtener la envolvente de cielo despejado, debido a que no es posible realizar el ajuste lineal robusto. No es posible, ya que hay días nublados en los que no hay ningún punto, o si los hay, no consiguen pasar el filtro para tomarlos como datos de DNI de cielo despejado.

Hasta ahora, cuando el algoritmo ha encontrado un día del cual no se ha podido calcular el ajuste lineal robusto, el programa lo ha obviado, calculando el siguiente día. Esto ha tenido como resultado que, del vector de datos empaquetados, no existan 365 días completos, o dicho de otra forma no existen 365 conjuntos de 1000 puntos que desempaquetar.

La posibilidad de tener varios días menos en la base de datos empaquetados no debería ser un problema a la hora de utilizar este método. En este caso contaríamos con menos posibilidades de combinación. A pesar de ello, se podría generar un año a partir de una base de datos de menos de 365 días.

El problema viene cuando aplicamos este método para generar el año sintético a partir de los datos horarios. En primer lugar, no es posible calcular una envolvente para esos días nublados en base horaria, pero se podría usar cualquier otra envolvente de los datos empaquetados. Aunque esto conllevaría un error muy elevado en los días generados, ya que como no se ha empaquetado ningún día nublado, esos días nublados en base horaria sólo se compararían con días despejados. Con esto en ningún caso obtendríamos días generados parecidos a los medidos.

Para solucionar esto, se presentan varias posibilidades:

- Utilizar la envolvente del día anterior al que no se puede calcular la envolvente. Pero en el caso que estamos estudiando, el año 2014, desde el día 1 de enero hasta el 4 de enero, se dieron días nublados, los cuales ningún día pasaba el filtro de 300 W/m^2 . Por lo que desarrollar un algoritmo que busque el día más cercano a esos días y escogiese esa envolvente sería algo complejo.
- Utilizar una envolvente global. Esto es, estudiar el año completo de igual modo que se ha estudiado cada día de la serie anual y escoger los mayores valores de la linealización de DNI y elevación solar, a continuación, se haría un ajuste lineal robusto, obteniéndose un valor de E_0 y β .

En este caso, si el algoritmo encuentra un día en el que no es posible realizar el ajuste lineal, en lugar de ajustar, sustituiría el resultado de la linealización con los valores de la envolvente global.

En este caso hemos optado por la segunda opción y a continuación, se procederá al cálculo de la envolvente global característica del año 2014.

Para el cálculo de la envolvente global, máxima para la serie de valores anual, vamos a partir de los datos minutales, y aplicaremos el método hasta el ajuste lineal robusto, como recordamos, para el ajuste lineal robusto aplicábamos la siguiente fórmula:

$$\ln\left(\frac{DNI}{\varepsilon}\right) = \ln(E_0) - \beta \cdot \left(\frac{1}{\sin(\alpha)}\right)$$

Donde ε corresponde al valor de la corrección distancia tierra-sol.

En este caso, calculábamos los valores de $\ln(DNI)$ y $\left(\frac{1}{\sin(\alpha)}\right)$ para cada día, en este caso se guardarán dichos valores de todo el año. Una vez hecho esto para los valores de $\left(\frac{1}{\sin(\alpha)}\right)$, vemos que estos valores se mueven entre 1 y 3, por ello a continuación vamos a realizar un filtro sobre los datos y nos quedaremos con los valores de $\ln(DNI)/\varepsilon$ de mayor valor, para ello recorreremos los valores de $\left(\frac{1}{\sin(\alpha)}\right)$, de 1 a 3, con un paso de 0.1 entre cada valor máximo.

Hay que tener en cuenta que de la fórmula anteriormente enunciada el valor de E_0 no es un valor fijo, es un valor que engloba ciertos coeficientes, como el coeficiente de corrección por distancia tierra-sol. Como la base de datos a analizar contiene valores de DNI de todos los años, es necesario modificar estos valores con este coeficiente. Una vez hecha esta aclaración, se procede al cálculo de la envolvente:

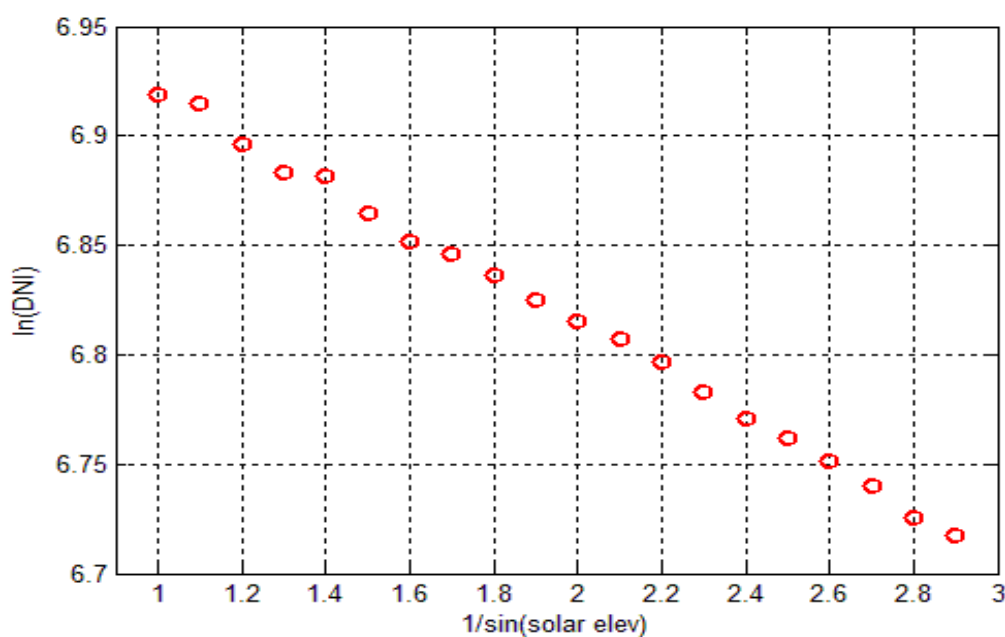


Figura 26. Valores máximos $\ln(DNI)$ para la serie anual 2014.

En la figura 26, se muestran los valores de $\ln(DNI)$ máximos de la serie anual de 2014, a estos valores se les hará el ajuste lineal robusto mediante el método 'Huber'.

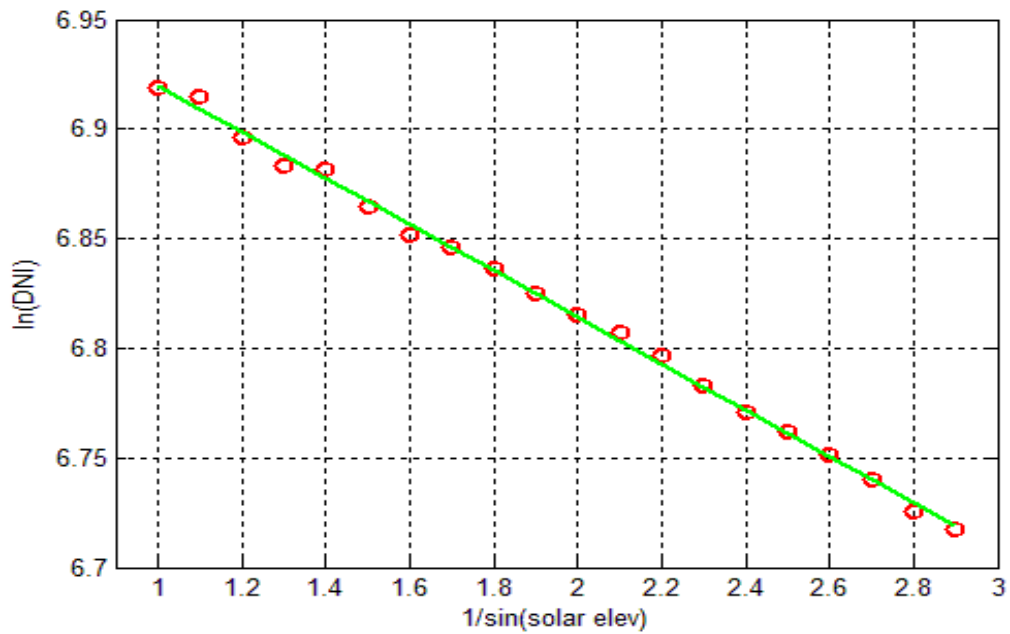


Figura 27. Ajuste lineal robusto envolvente global.

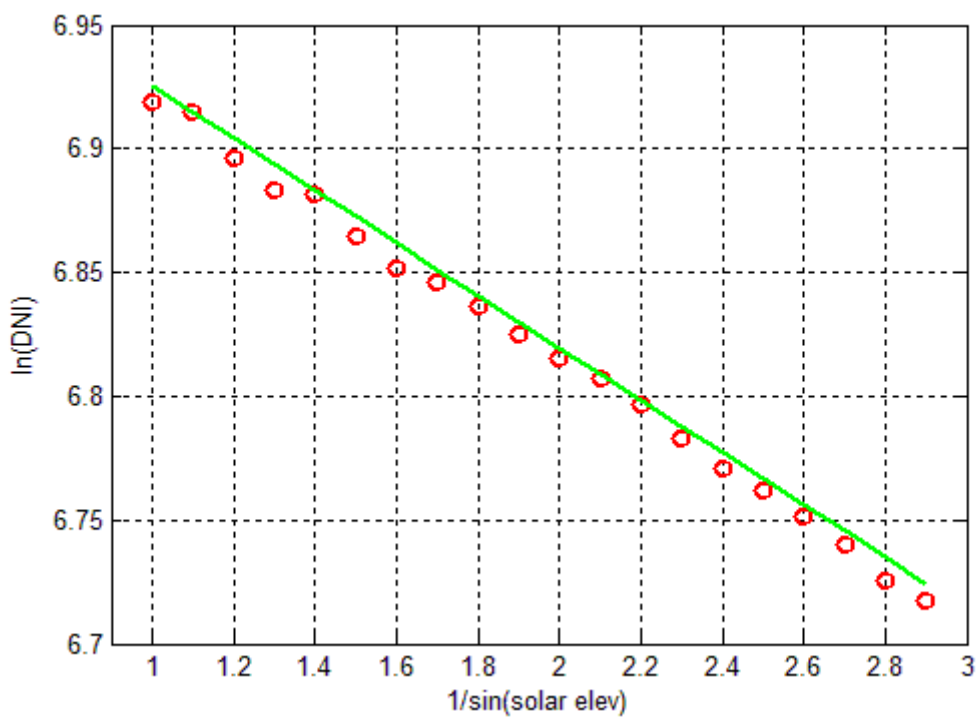


Figura 28. Ajuste lineal robusto envolvente global de máximo valor.

En la figura 27 se ha realizado el ajuste lineal, y en la figura 28 se ha realizado el ajuste para que la envolvente pase por encima del valor máximo de DNI de la serie anual. Este ajuste nos lleva a unos valores de envolvente global de $E_0=1125.29$ y $\beta=0.1059$. Hay que tener en cuenta el valor del coeficiente distancia tierra sol antes de aplicar esta envolvente a un día determinado.

6.4 GENERACIÓN DE AÑOS SINTÉTICOS

Una vez calculados los valores de E_0 y β , que corresponden a la envolvente global, tenemos una curva de cielo despejado que es máxima independientemente del día del año.

El método de C.M. Fernández Peruchena, propone que con los valores de E_0 y β de toda la serie anual, mediante combinaciones entre dichos valores y toda la serie de días empaquetado, el algoritmo busque el día más cercano en términos de distancia euclídea.

Probar para cada día una combinación de E_0 y β distinta partiendo de una base de datos de más de 8.000 valores, conllevaría un altísimo coste combinatorial. Por ello la serie anual sintética que vamos a generar sólo tendrá en cuenta dos pares de valores de E_0 y β . El primer par será el correspondiente a la envolvente calculada con los datos horarios y el segundo par corresponderá a los datos del empaquetado.

Este método después será comparado mediante un estudio estadístico con otro método de generación de series de DNI.

6.4.1 Generación año 2014

El método utilizado para generar este año consiste en calcular la envolvente tanto para los datos diezminutales, los cuales van a ser empaquetados y por lo cual obtendremos una serie de valores de E_0 y β para la serie de años de 2000 a 2012, como para los datos horarios, usando también el algoritmo de aumento a escala minutal.

El uso de la envolvente global se limitará a casos en los que el algoritmo no sea capaz de calcular unos valores de E_0 y β propios.

Si se estudia con detenimiento este método, lo que estamos obteniendo es, una base de datos y unos valores de empaquetamiento para los días despejados y otra base de datos para los días nublados. Es decir, el algoritmo de comparación entre los días y el empaquetamiento, va a comparar días despejados con días despejados y días nublados con días nublados.

Consideraremos un día nublado aquel que no cuente con suficientes puntos como para realizar un ajuste lineal robusto, es decir, no utilizaremos el concepto de la k_b .

El algoritmo se basará en datos empaquetados de DNI diez minutales, con sus respectivos valores de E_0 y β , menos en los días nublados que usará los valores de E_0 y β de la envolvente global.

En el caso de los datos DNI horarios, la envolvente de cielo despejado será calculada para cada día con la ampliación minutal, y en el caso de no poder realizarse el ajuste lineal, impondrá los valores de E_0 y β de la envolvente global.

El algoritmo, para buscar el día más parecido desempaquetará utilizando dos envolventes distintas, la primera la calculada con los datos horarios y la segunda, la propia del día que se ha

empaquetado previamente y mediante una comparación punto a punto, elegirá el día desempaquetado más parecido al medido.

Por lo que una vez aplicado este método, tendremos un año 2014 generado a partir de días de los años comprendidos entre 2000 y 2012 en escala diezminutal.

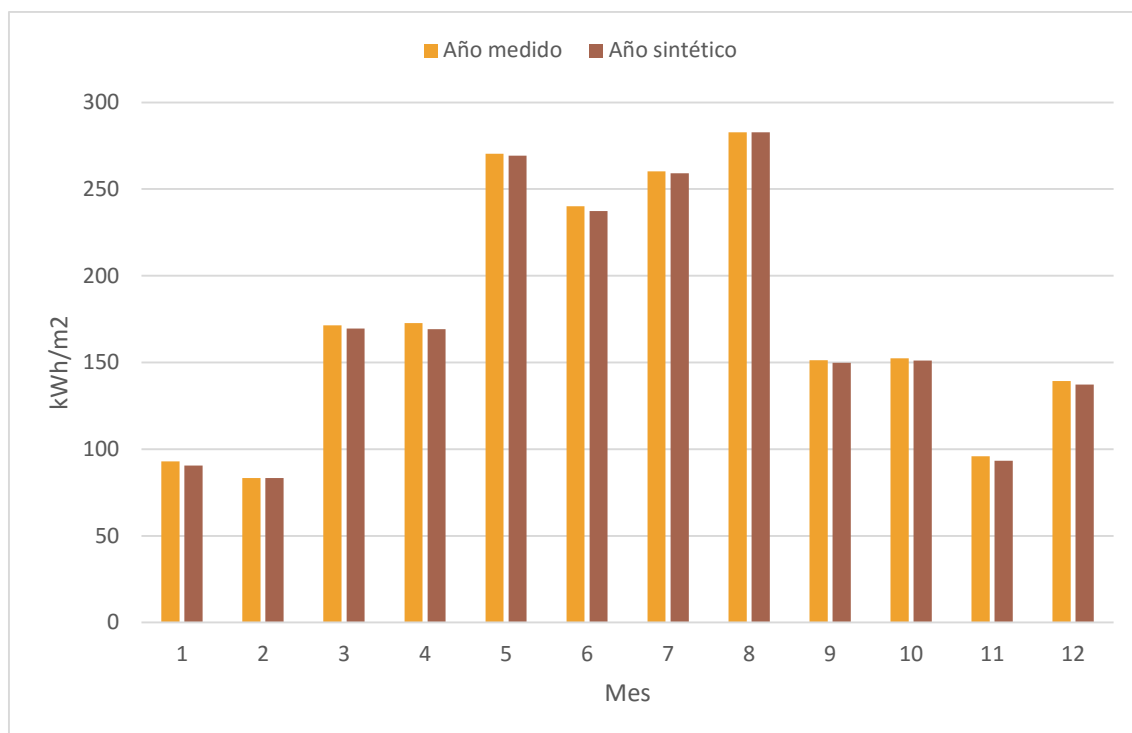


Figura 29. Comparación valor acumulado mensual.

En la figura 29, podemos observar el valor acumulado mensual tanto para el año medido, como para el año generado. A primera vista podemos advertir diferencias entre los valores acumulados, pero esta diferencia no es considerable. Para ver esto, a continuación, se muestra una tabla con los errores relativos cometidos entre los valores acumulados mensuales medidos y los generados:

Mes	1	2	3	4	5	6	7	8	9	10	11	12
Error (%)	2.45	0.54	0.27	1.93	0.34	1.08	0.44	0.04	1.46	0.92	1.51	1.63

Tabla 3. Error acumulado mensual año sintético respecto año medido.

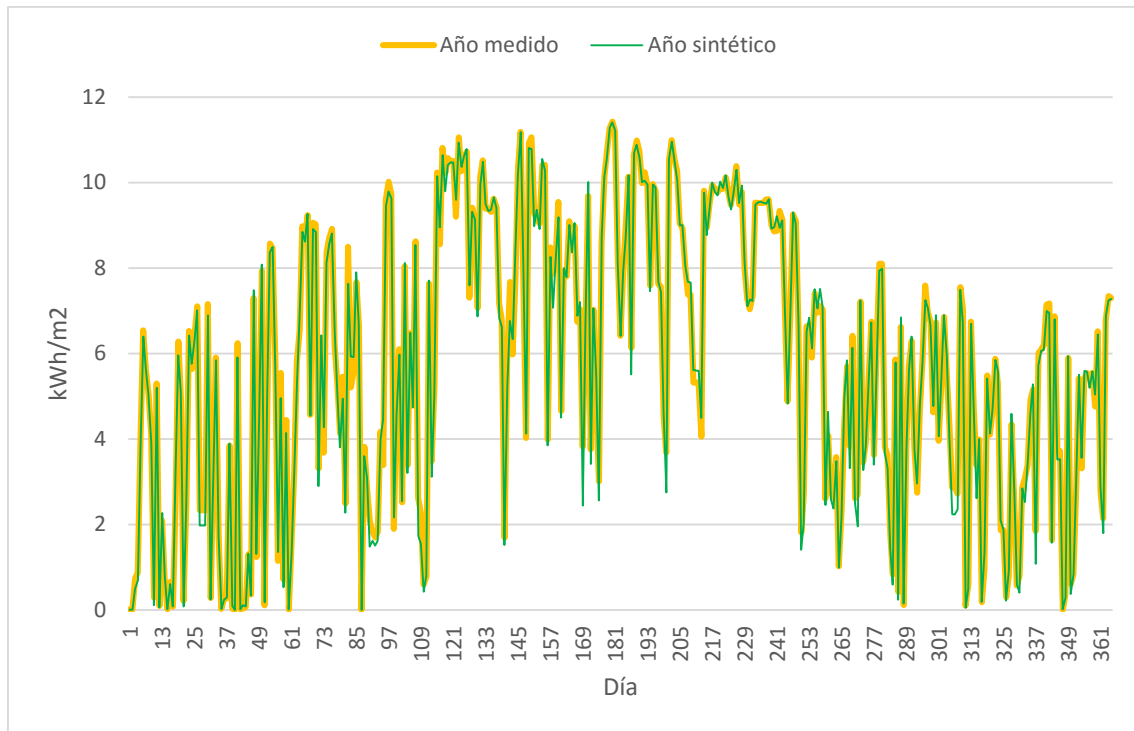


Figura 30. Comparación valor acumulado diario.

En la figura 30, se muestran los valores diarios acumulados en kWh/m^2 . A la vista del gráfico vemos como los días generados sintéticamente siguen la tendencia del día medido. Pero para compararlos correctamente calcularemos en error relativo cometido:

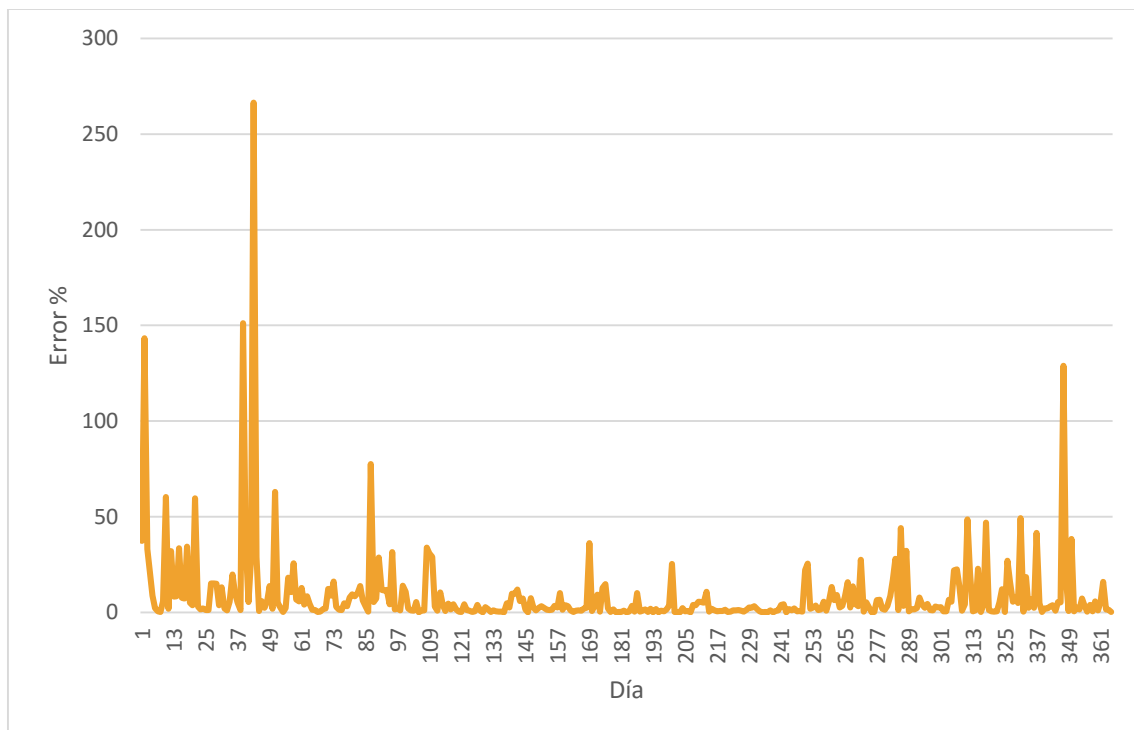


Figura 31. Error diario año sintético respecto a año medido.

En la figura 31, podemos observar el error porcentual cometido entre los valores acumulados diarios correspondientes al año medido y al año sintético. A la vista de los resultados, vemos días en los que los errores son muy elevados, llegando a superar el 250% de error. Por ello vamos a analizar esos días, en los que el error relativo supera el 100% en los valores acumulados diarios.

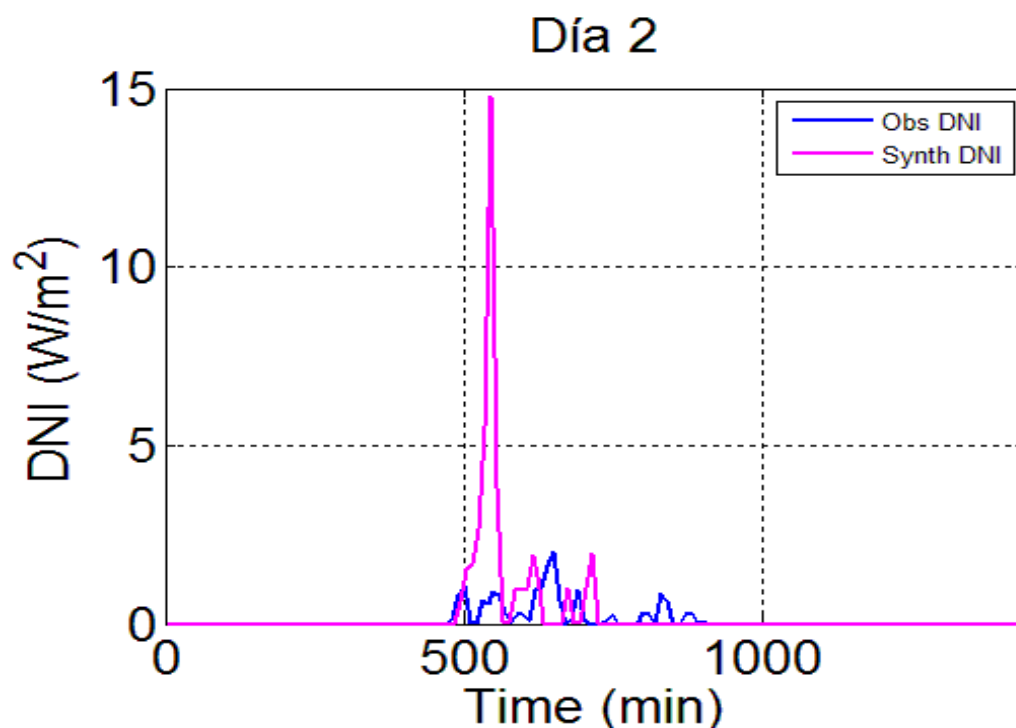


Figura 32. Día 2 enero de 2014 desempquetado con día 327 de 2000. Error 143%.

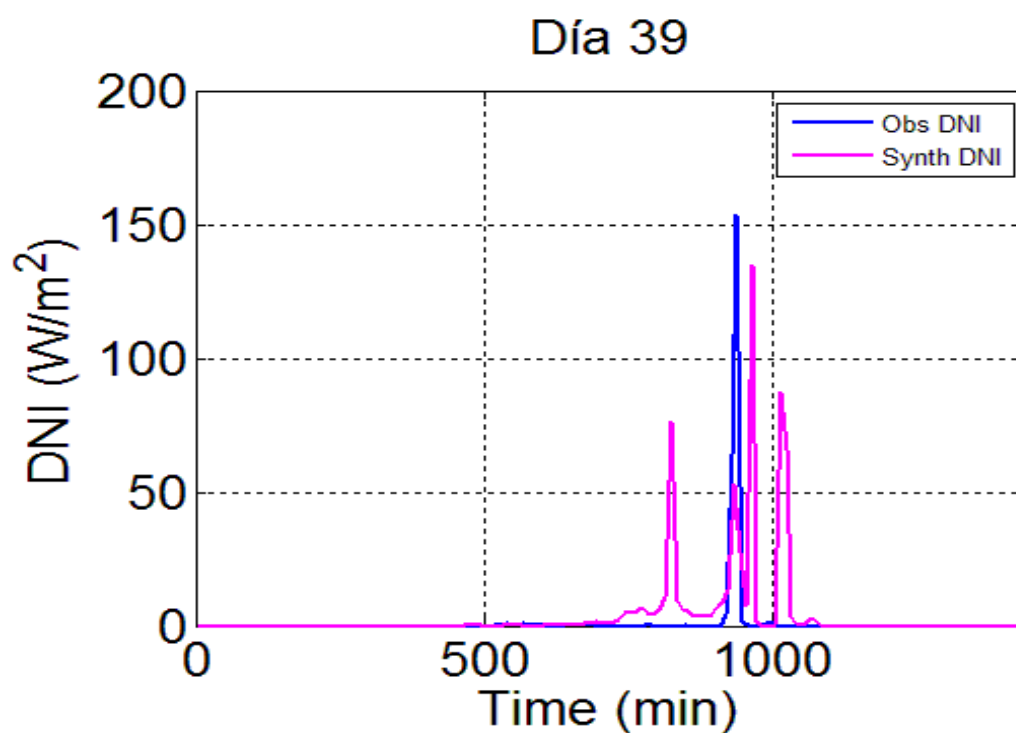


Figura 33. Día 8 febrero de 2014 desempquetado con día 66 de 2001. Error 151%.

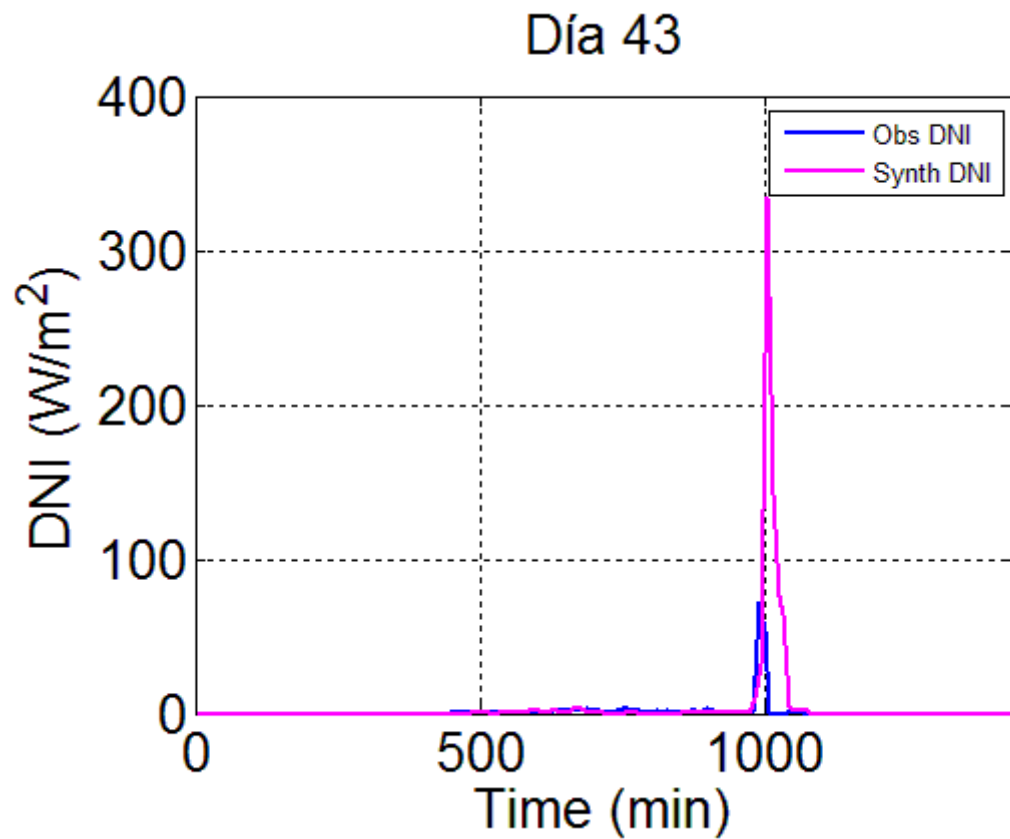


Figura 34. Día 12 febrero de 2014 desempaquetado con día 340 de 2003. Error 266%.

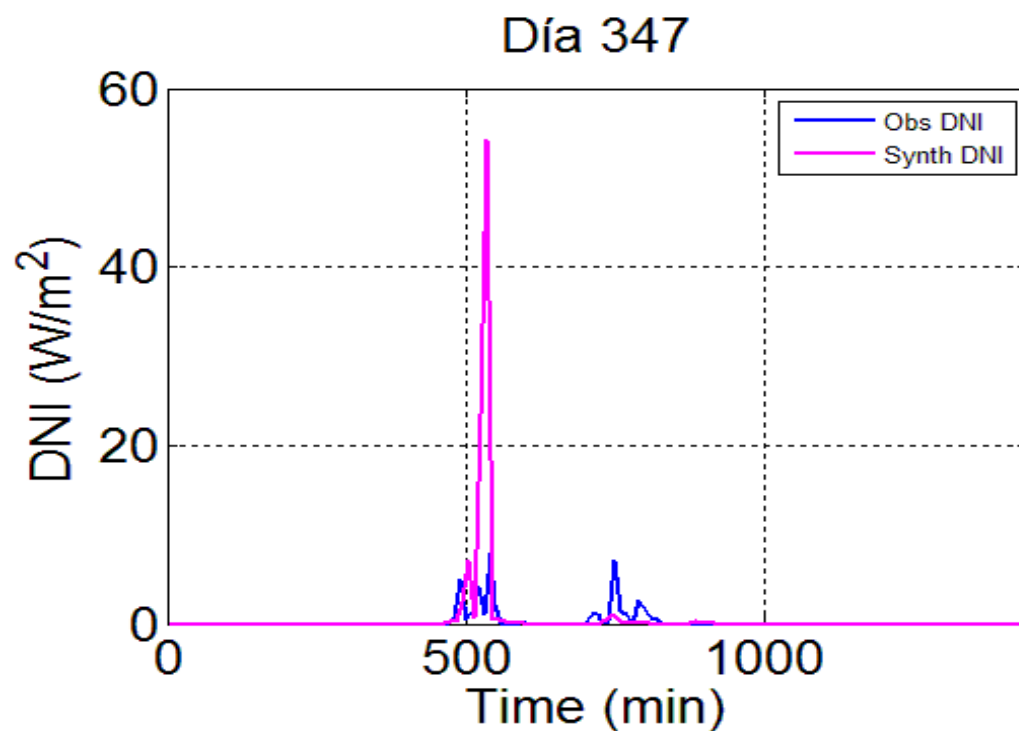


Figura 35. Día 13 diciembre de 2014 desempaquetado con día 67 de 2011. Error 128%.

Desde la figura 32 a la figura 35 podemos ver los días en los que se comete mayor error respecto al valor de DNI acumulado. Como vemos en todos los casos se tratan de días nublados. Al ser nublados todos utilizan la envolvente global, lo que limita el número de días con los que pueden compararse. Pero el principal problema por el cual el error es tan elevado es la forma de buscar el día más parecido, ya que solo usa los datos de DNI horarios. Esto puede hacer que, por ejemplo, en la figura 34 veamos cómo pueden coincidir los datos horarios, pero como el resto no se comparan, sale un pico de DNI no esperado. Este problema se ve atenuado debido a que también comparamos respecto a valores acumulados parecidos.

Aunque estos problemas en ciertos días nublados no tienen tanta importancia ya que, como vemos en la figura 32, hay días con muy poca irradiancia, menor de 15 W/m^2 . En los que una pequeña diferencia, conlleva un error elevado.

Ahora, una vez analizado esto, eliminamos estos valores de la figura 31, para ver con mayor resolución el resto de los errores:

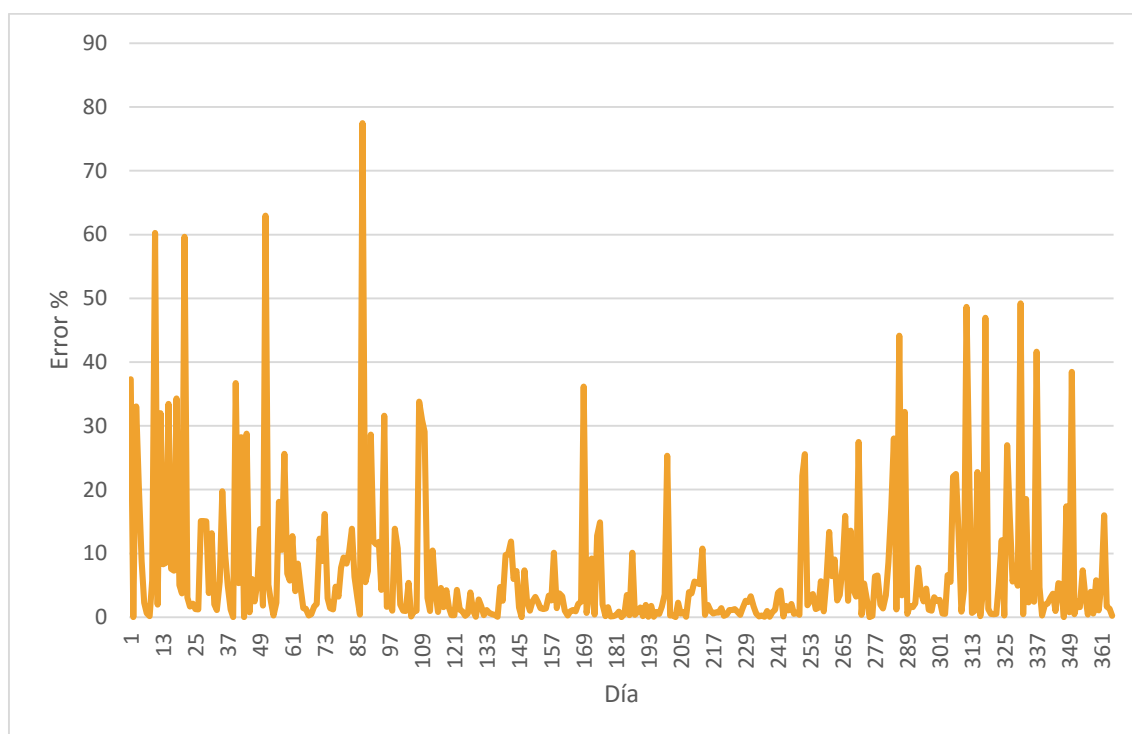


Figura 36. Error diario año sintético respecto a año medido. Sin errores extremos mayores de 100%

De igual modo existen 4 días, como vemos en la figura 36, que sobrepasan un error del 50%, y que procedemos a analizar.

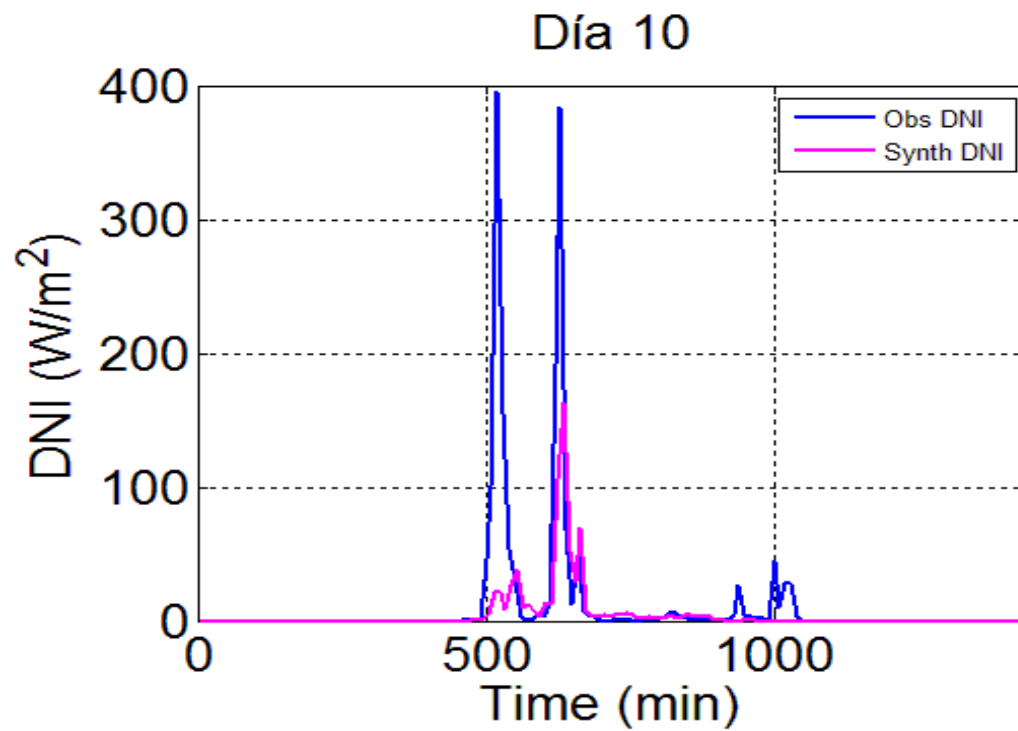


Figura 37. Día 10 enero de 2014 desempquetado con día 34 de 2002. Error 60%.

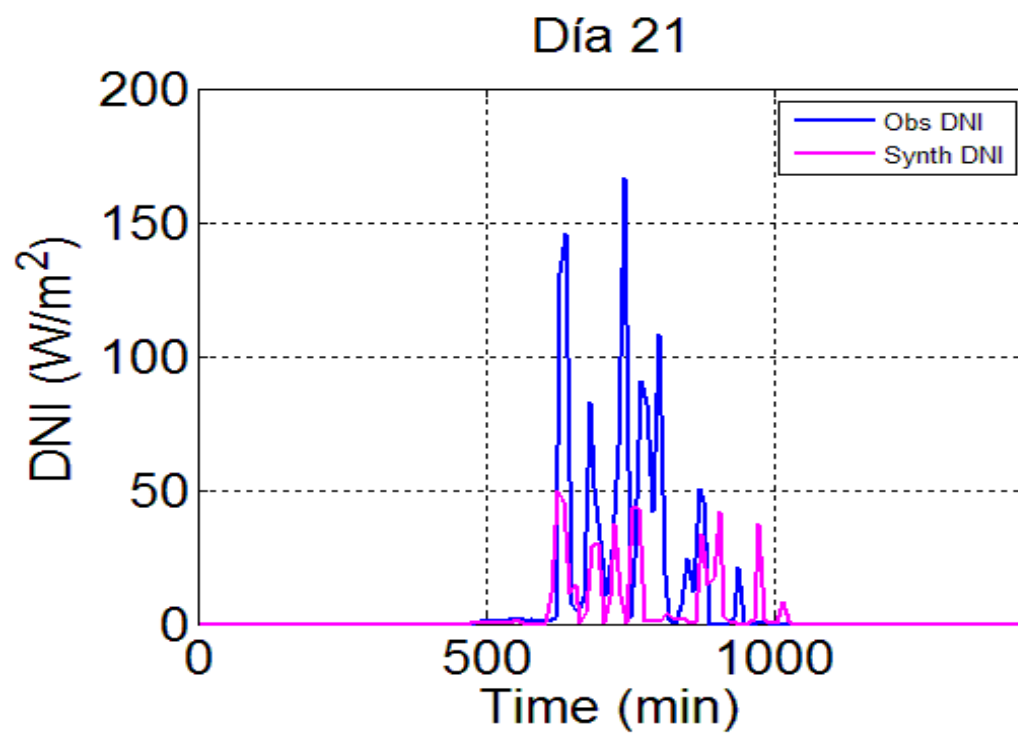


Figura 38. Día 21 enero de 2014 desempquetado con día 104 de 2000. Error 59%.

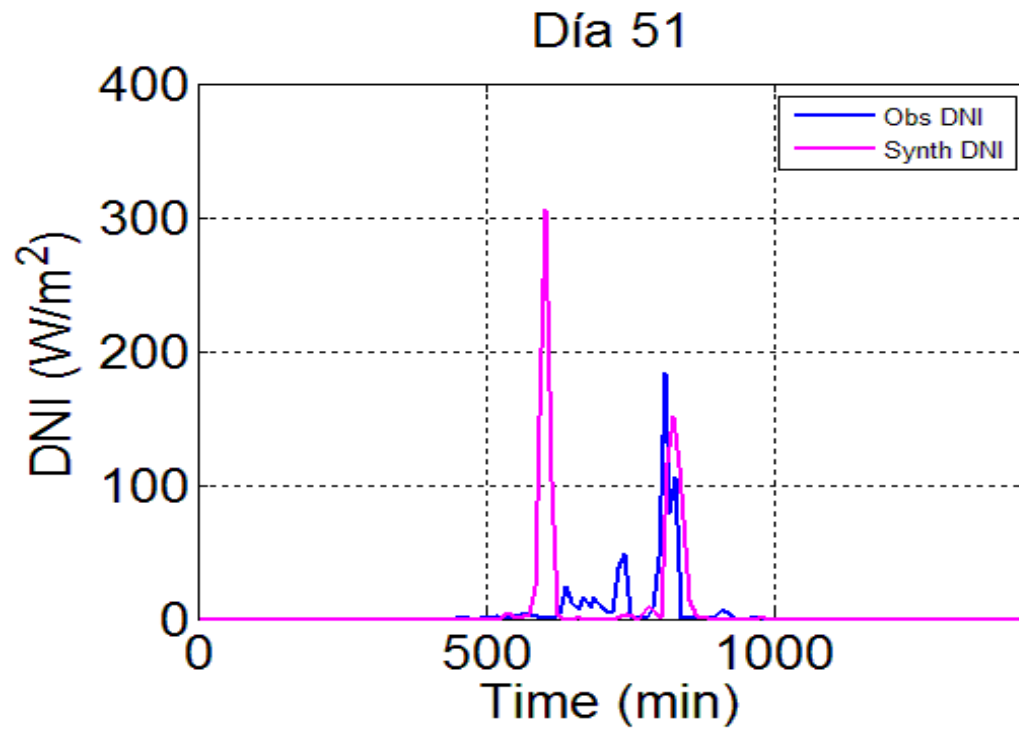


Figura 39. Día 20 febrero de 2014 desempquetado con día 61 de 2002. Error 63%.

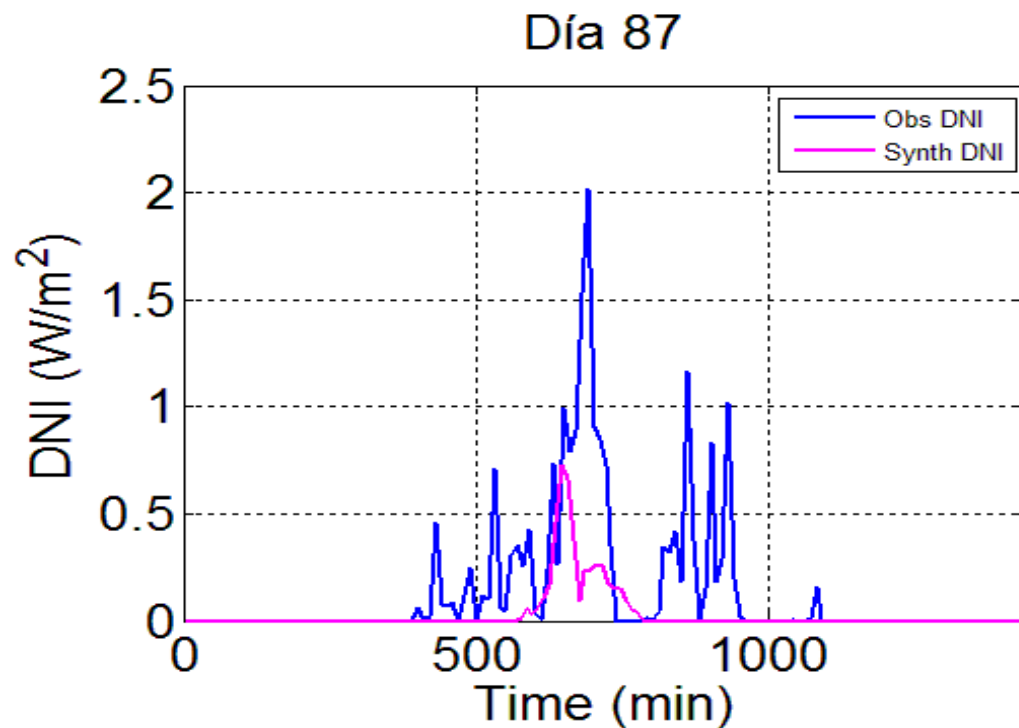


Figura 40. Día 28 marzo de 2014 desempquetado con día 324 de 2011. Error 77%.

Como observamos desde la figura 37 a la figura 40, se tratan de nuevo de días nublados. En este caso el de mayor error es el día 87 con un 77% de error en el acumulado diario, pero como observamos en la figura 40, aunque el error porcentual sea alto, los valores de DNI son muy

bajos como para tener importancia. Una vez vistos los errores, nos quedamos de nuevo con los errores diarios acumulados:

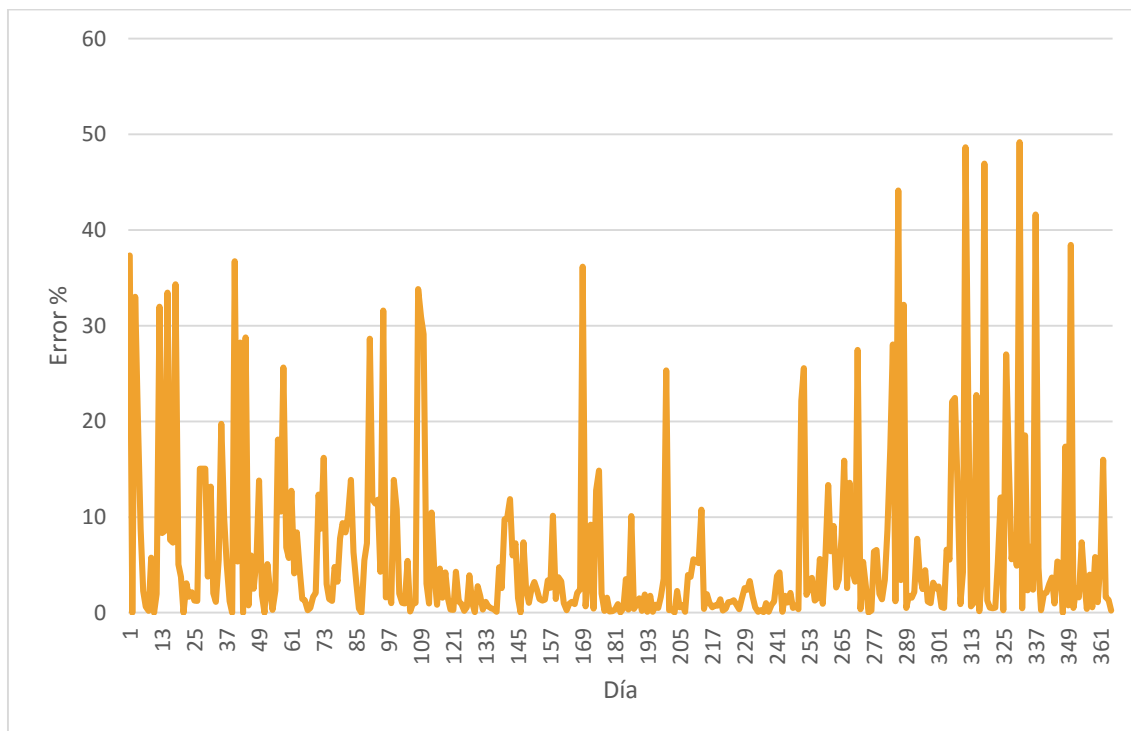


Figura 41. Error diario año sintético respecto a año medido. Sin errores extremos mayores de 50%.

7 ANÁLISIS ESTADÍSTICO DE MODELOS

Una vez aplicado el método de Fernández Peruchena [1] en la base radiométrica del GTER, hemos obtenido un año sintético generado a partir de una base de datos diezminutal de los años 2000 al 2012, y a partir de datos horarios de 2014.

De la misma base de datos contamos con el año 2014 en base diezminutal generado mediante el método presentado por Larrañeta [3]. Estos años sintéticos van a ser comparados entre sí, ya que todos generan datos diezminutales a partir de datos horarios partiendo de la misma base:

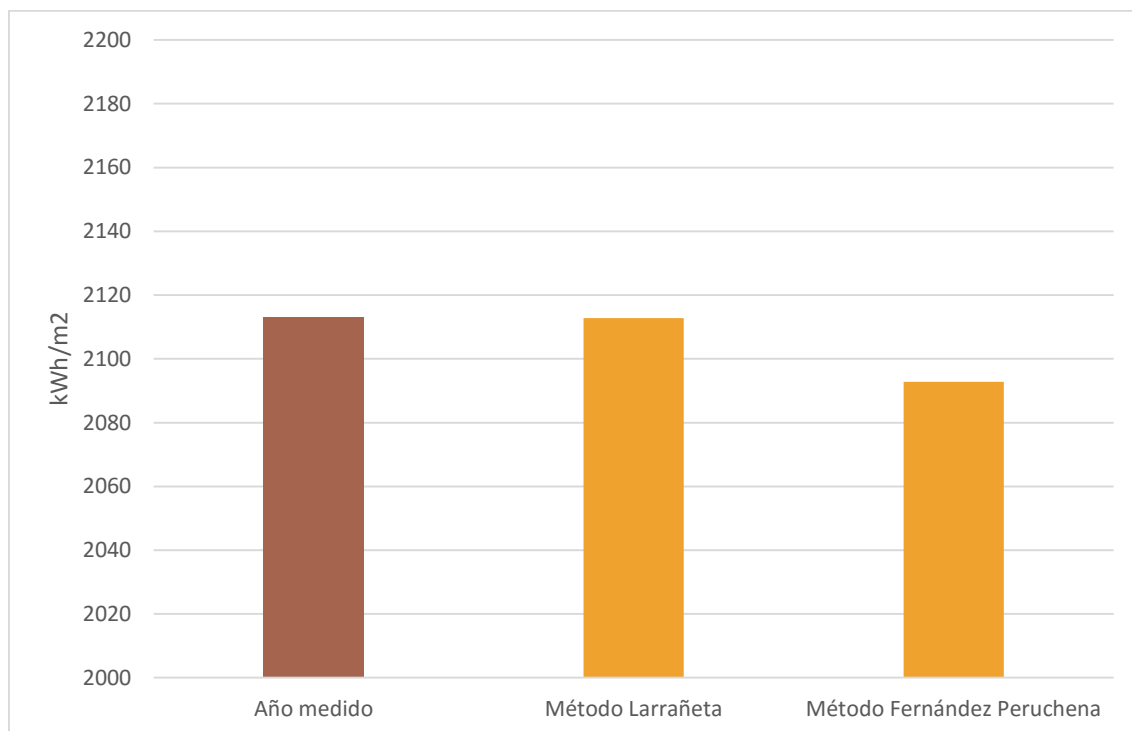


Figura 42. Comparación datos acumulados anuales entre años generados 2014.

	Año medido	Método Larrañeta	Método Fernández Peruchena
Valor acumulado 2014 (kWh/m^2)	2113.06	2112.75	2092.77
Error (%)	-	0.014	0.959

Tabla 6. Valores acumulados año 2014.

Como podemos observar tanto en la figura 42 como en la tabla 6, el año generado con el método Larrañeta, con el que vamos a comparar, es el que más se aproxima en términos de valor

acumulado. Mientras que el año generado sintéticamente con el método de Fernández Peruchena, se obtiene un valor acumulado menor.

Debido a esto, como vemos en la tabla, el método de Larrañeta obtiene un error del 0.014% y el método de Fernández Peruchena, un error del 0.959%, valor muy cercano al 1%, ya que tiene 20.29 kWh/m^2 menos.

A pesar de ello, esto no demuestra que un método sea mejor que otro, ya que existen otro tipo de índices también importantes, como puede ser la distribución de frecuencia o el valor de KSI, que analizaremos a continuación para cada método.

El índice KSI es un indicador de similitud de distribución entre dos series de valores. En este caso nos mostrará la similitud entre las dos series de datos, la serie generada y la serie medida.

KSI, test integral Kolmogorov-Smirnov, se define como la diferencia integral entre las dos funciones de distribución (CDF, Cumulative Distribution Function). El KSI se define por la siguiente integral:

$$KSI = \int_{x_{max}}^{x_{min}} D_n dx$$

Siendo D_n una variable discreta y el número de intervalos integrados idéntico en todos los casos, esta integración trapezoidal es usada sobre todo el rango de la variable x independiente. El valor relativo de KSI, el KSI porcentual, es obtenido mediante la normalización del área crítica:

$$KSI = 100 \cdot \frac{\int_{x_{max}}^{x_{min}} D_n dx}{a_{crítico}}$$

Donde $a_{crítico}$ es calculado como:

$$a_{crítico} = V_c \cdot (x_{max} - x_{min})$$

El término V_c puede ser determinado por el valor N , que es el número de datos de los que se parte:

$$V_c = \frac{1.63}{\sqrt{N}}, \quad N \geq 35$$

El mínimo valor del índice KSI porcentual es cero, que significa que las dos funciones de distribución son exactamente iguales, dicho esto, un valor de este índice menor a 100% cuando ambas funciones de distribución son cercanas entre ellas, para nuestro objetivo, un valor de KSI (%) por debajo de 100% es considerado como una prueba de que el año sintético es estadísticamente indistinguible del año de medidas reales de valores diez minutales.

Una vez definido este índice procederemos al análisis del valor del índice KSI (%) de cada método.

7.1 ANÁLISIS KSI (%)

Para analizar el valor de KSI (%) y la función de distribución, hacemos uso de dos representaciones gráficas, como mostramos en la figura 43 y 44.

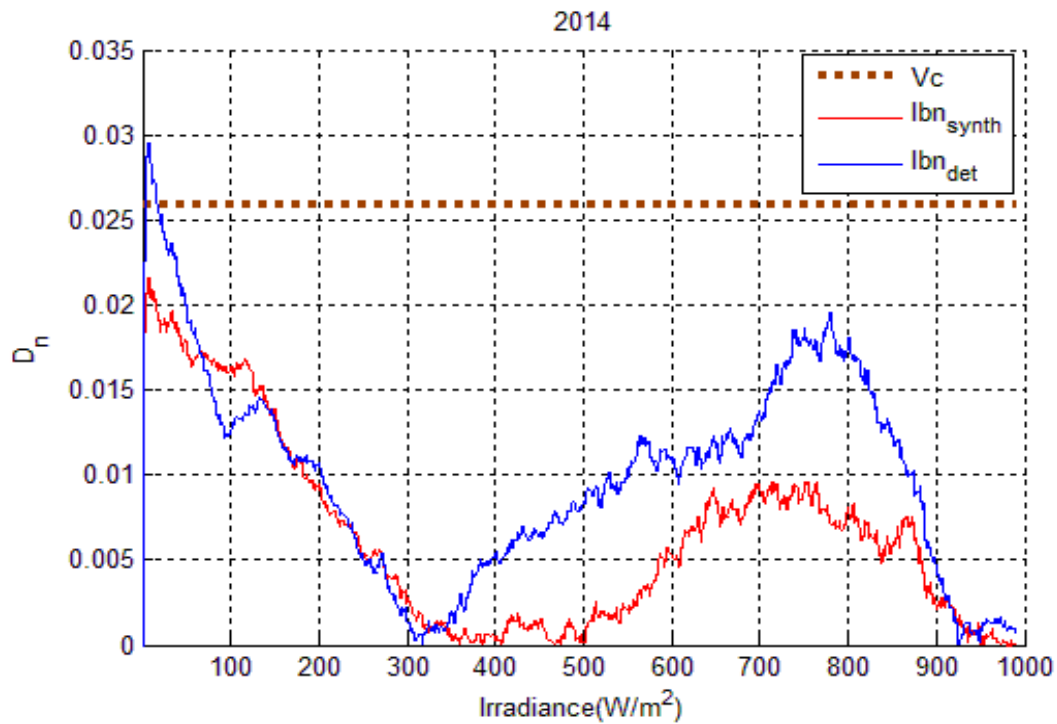


Figura 43. Límite tolerable valor D_n . Método Larrañeta (Ibn_{det}) y Método 1 (Ibn_{synth}).

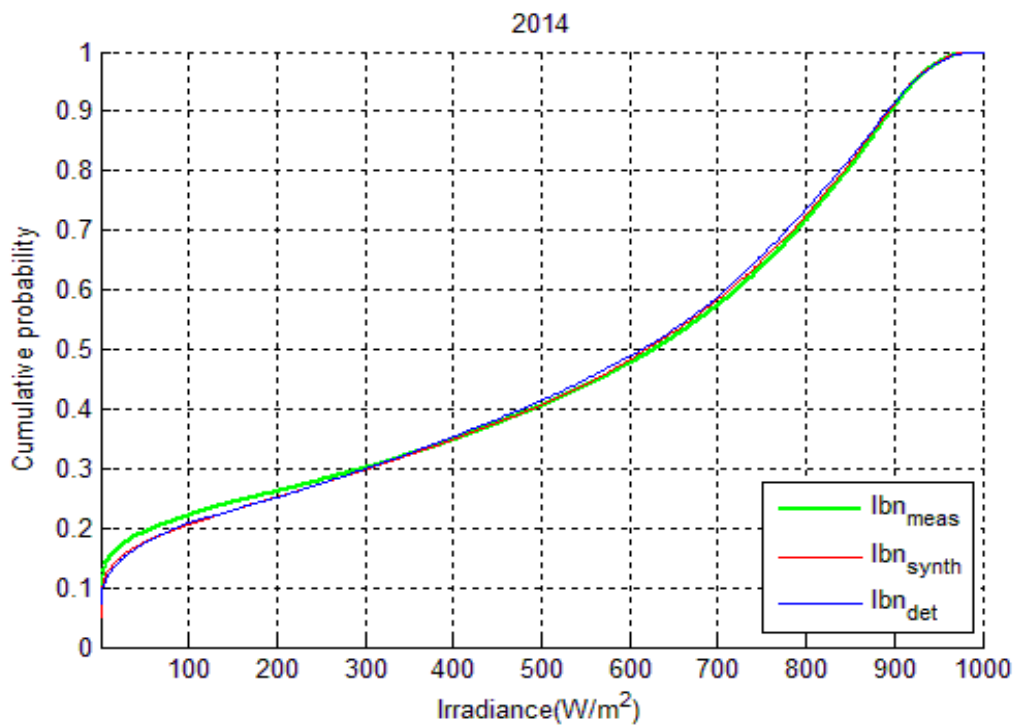


Figura 44. Función de distribución. Método Larrañeta (Ibn_{det}) y Método 1 (Ibn_{synth}).

En cada una de ellas se muestran dos curvas, una azul y otra roja. La azul representa la curva perteneciente al año generado mediante el método Larrañeta y la curva roja representa al año sintético generado mediante el método de Fernández Peruchena.

En la figura 43, además de las dos curvas observamos una línea punteada horizontal, que representa el límite del valor D_n , si este límite es sobrepasado, mostraría diferencias importantes entre el modelo generado y el año medido, como podemos observar, el modelo sintético generado por el método de Fernández Peruchena no sobrepasa esta línea, por lo que nos encontramos ante un modelo muy aproximado al real, pero esto es representado por el valor de KSI (%).

El algoritmo que calcula dicho valor nos muestra un resultado de KSI sintético=24.87 %, mientras que el modelo del año generado por Larrañeta muestra un resultado de KSI=38.43 %, por lo que el modelo generado por el método de Fernández Peruchena obtiene un valor de KSI menor que el obtenido por Larrañeta, pero ambos están por debajo de un valor KSI de 100% por lo que son

7.2 OTROS INDICADORES ESTADÍSTICOS

Una vez analizado la similitud en las funciones de distribución, aún existen otros factores utilizados para comparar series de datos, como pueden los indicadores de dispersión y los indicadores de rendimiento global.

Por ello vamos a analizar estos distintos factores con el objetivo de elegir el método que más se adecua al objetivo inicial.

Previo a la utilización de estos indicadores, hay que definir los distintos factores que se utilizan para el cálculo de estos índices estadísticos. Distinguiremos con la letra O, los valores observados o medidos, y con la letra S, los valores sintéticos o generados por nuestros modelos:

- N se refiere al número de pares de valores de los cuales tenemos datos, hay que destacar que en este análisis sólo se han tenido en cuenta valores medidos en las horas con luz del día, para lo cual hemos tenido en cuenta datos cuyos valores temporales corresponden a elevaciones solares por encima de 6°.
- O_m y S_m corresponden a los valores medios de la serie de valores de la serie de datos medidos y la serie de datos sintética respectivamente.
- o_i y s_i corresponden a los valores de cada par i, de los datos medidos y los datos sintéticos.
- o_{max} y o_{min} corresponden al valor máximo y valor mínimo de la serie medida de datos.

7.2.1 Indicadores de dispersión

Para el indicador de dispersión se utilizará el llamado RMSD (Root mean squared difference), que es una medida que representa la derivación estándar de las diferencias entre los valores predichos y los observados. Este valor es calculado mediante la siguiente expresión:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (o_i - p_i)^2}$$

Para la comparación de los distintos modelos haremos uso de este índice, pero con cierta modificación, lo que nos permitirá analizar este valor en término porcentual referido a O_m , por ello utilizaremos el llamado NRMSD (Normalized root mean squared difference), que representa el valor normalizado de RMSD y tiene la siguiente expresión:

$$NRMSD = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (o_i - p_i)^2}}{o_{max} - o_{min}}$$

Donde el valor de $o_{max}=993.23 \text{ W/m}^2$ y $o_{min}=0 \text{ W/m}^2$. Y el número de pares de valores es de $N=24022$.

Definidos estos valores el resultado se muestra en la siguiente tabla:

	Método Larrañeta	Método Fernández Peruchena
NRMSD (%)	12.36	13.63

Tabla 7. Comparación NRMSD años sintéticos año 2014.

Como se trata de una indicación de dispersión, un valor más bajo, conlleva una dispersión menor respecto al año medido, por lo que, a menor valor, más parecido es el año sintético. Como vemos en la tabla 7, el año sintético con menor índice de dispersión, es el año generado con el método Larrañeta, siendo el que muestra mayor dispersión el año generado con el método de Fernández Peruchena, aunque la diferencia entre ellos es menor de 2 puntos.

7.2.2 Indicadores de rendimiento

Otro indicador útil para la comparación de bases de datos generados respecto a datos medidos son los indicadores de rendimiento, en este caso utilizaremos 3 indicadores distintos:

- Nash-Sutcliffe efficiency (NSE). Este índice tiene un rango desde $-\infty$ a 1. Si el valor es igual a 1, significa que el modelo es exacto a la serie de datos medidos, si el valor es 0, indica que las predicciones son tan precisas como lo es la media de los datos observados, en cambio, si el valor es menor que 0, indica que la variabilidad de los datos generados es mayor que la de los datos reales, por lo que no se considera un buen modelo. Se define mediante la siguiente expresión:

$$NSE = 1 - \left[\sum_{i=1}^N (s_i - o_i)^2 \right] / \left[\sum_{i=1}^N (s_i - o_m)^2 \right]$$

- Willmott index of agreement (WIA). Este índice varía entre los valores 0 y 1. Este índice detecta las diferencias acumulativas y proporcionales entre los valores generados y medidos, y es extremadamente sensible a valores extremos debido a sus diferencias cuadráticas. El valor 0 correspondería a una concordancia nula y el valor 1 correspondería a una coincidencia absoluta. Se define mediante la siguiente expresión:

$$WIA = 1 - \left[\sum_{i=1}^N (s_i - o_i)^2 \right] / \left[\sum_{i=1}^N (|s_i - O_m| + |o_i - O_m|)^2 \right]$$

- Legates coefficient of efficiency (LCE). Índice con un rango desde $-\infty$ a 1. Siendo 1 concordancia absoluta y los valores menores que 0, diferencias importantes entre el modelo y los valores reales. Se define con la siguiente expresión:

$$LCE = 1 - \left[\sum_{i=1}^N |s_i - o_i| \right] / \left[\sum_{i=1}^N |o_i - O_m| \right]$$

Una vez definidos estos índices de rendimientos se procede a su cálculo para cada uno de los modelos. Obteniéndose el resultado mostrado en las siguientes tablas:

	Método Larrañeta	Método Fernández Peruchena
NSE	0.8665	0.8377

Tabla 8. Comparación NSE años sintéticos año 2014.

	Método Larrañeta	Método Fernández Peruchena
WIA	0.9651	0.9573

Tabla 9. Comparación WIA años sintéticos año 2014.

	Método Larrañeta	Método Fernández Peruchena
LCE	0.7703	0.7579

Tabla 10. Comparación LCE años sintéticos año 2014.

En los tres valores analizados, la mayor aproximación a 1 supone una mayor coincidencia estadística entre los valores medidos y los valores sintéticos generados, lo que nos hace destacar el método de Larrañeta por encima del de Fernández Peruchena, aunque hay que decir que son valores muy próximos.

8 CONCLUSIONES

La aplicación de esta metodología ha presentado dificultades que procedemos a analizar.

La más importante es la imposibilidad de empaquetar los días nublados, que hemos intentado solucionar mediante el uso de la envolvente global máxima. Y respecto al cálculo de la envolvente a partir de datos horarios, hemos introducido la ampliación de frecuencia mediante una interpolación lineal.

Otra dificultad en el caso del algoritmo, es el desempaquetado de días utilizando todos los valores disponibles de E_0 y β . Fernández Peruchena propone en el desempaquetado, la utilización de todos los valores de E_0 y β calculados durante el empaquetado.

En nuestro caso hemos empaquetado 13 años, pero si sólo hubiésemos empaquetado 1, la aplicación del método tal como se propone, supondría la utilización de 365 E_0 y 365 β . Combinados entre sí, darían lugar a más de 133.000 posibilidades. Es decir, para cada día habría que comparar los datos horarios con el desempaquetado que dan lugar esas 133.000 posibilidades, y eso para sólo 1 día. Por ello, hemos optado por utilizar sólo la envolvente calculada a partir de los datos horarios y la propia del día empaquetado

Teniendo en cuenta sólo estas 2 combinaciones, el algoritmo puede tardar horas, lo que supondría un altísimo coste computacional. Añadiendo a esto que no es sólo 1 año, sino 13. A pesar de ello podemos ver cómo estas dos combinaciones han generado un año con resultados estadísticos aceptables. Aunque estos resultados se pueden perfeccionar si introducimos mejoras. Una de estas mejoras sería cambiar la forma de calcular la envolvente, ya que como dijimos en el punto 6.1, al necesitar el ajuste robusto un mínimo de puntos para ajustar la envolvente, sería aconsejable estudiar cuál es el número mínimo de puntos para que el cálculo sea el óptimo.

A la vista de los resultados de las tres tablas. Tabla 8, tabla 9 y tabla 10, que representan los valores de NSE, WIA y LCE respectivamente, podemos concluir que, estadísticamente, el año sintético generado mediante el método de Larrañeta, es el que más se ajusta al año base 2014 de la base de datos del GTER.

A pesar de que todos estos índices muestran la similitud entre los valores generados y los valores medidos, los resultados del estudio de la función de distribución se postula como clave y como hemos visto en el análisis del punto 7, tanto el método de Larrañeta como el método de Fernández Peruchena, se muestran estadísticamente indiferenciables respecto al año medido. El modelo presentado con el método Larrañeta, ha mostrado también valores muy positivos en los índices estadísticos, en cambio, en el análisis del KSI, vimos como para ciertos valores sobrepasa el límite, y presenta un KSI (%), mayor que el del método de Fernández Peruchena.

Por lo tanto, podemos concluir que en términos de distribución el método de Fernández Peruchena es más acertado, pero el de Larrañeta obtiene un valor de distribución también ajustado y obtiene mejores resultados en el análisis de índices estadísticos.

Puede ser consultado en el Anexo I, los días y años a partir de los cuales se han generado cada uno de los días de 2014 y el año generado por el método de Fernández Peruchena puede ser consultado en el Anexo II. En el que se muestran las curvas con los valores diez minutales tanto sintéticos como medidos para todo el año 2014.

9 REFERENCIAS

- [1] Carlos M. Fernández-Peruchena, Manuel Blanco, Martín Gastón, Ana Bernardos. Increasing the temporal resolution of direct normal solar irradiance series in different climatic zones. *Solar Energy* 115 (2015) 255–263.

- [2] Carlos M. Fernández-Peruchena, Manuel Blanco, Ana Bernardos. Generation of series of high frequency DNI years consistent with annual and monthly long-term averages using measured DNI data. *Energy Procedia* 49 (2014) 2321 – 2329. SolarPACES 2013.

- [3] Miguel Larrañeta, Sara Moreno-Tejera, Manuel Antonio Silva-Pérez, Isidoro Lillo-Bravo. An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series. *Solar Energy* 122 (2015) 517–528.

- [4] Valeriano Ruiz Hernández, Manuel A. Silva Pérez. Radiación Solar. Grupo de Termodinámica y Energías Renovables, Escuela Superior de Ingenieros, Universidad de Sevilla.

- [5] Gall, J., Abel, D., Ahlbrink, N., Pitz-Paal, R., Anderson, J., Diehl, M., Teixeira Boura, M., Schmitz, M., Hoffschmidt, B., 2010. Simulation and control of thermal power plants. In: *Proceedings of the International Conference on Renewable Energies and Power Quality (ICREPQ'10)*, Granada (Spain), March 23–25, 2010, pp. 294–298.

- [6] Manuel A. Silva Pérez, Sara Moreno Tejera. Radiación solar. *Apuntes 4º Grado Ingeniería de la Energía* 2015-2016.

10 ANEXO I

Día 2014	Día emp	Año						
1	328	2010	41	244	2012	83	335	2007
2	327	2000	42	54	2004	84	363	2002
3	231	2005	43	340	2003	85	55	2006
4	5	2003	44	124	2004	86	136	2009
5	240	2002	45	138	2008	87	324	2011
6	285	2006	46	114	2001	88	298	2004
7	265	2003	47	203	2007	89	5	2002
8	121	2010	48	143	2007	90	98	2000
9	48	2011	49	142	2000	91	294	2001
10	34	2002	50	52	2012	92	360	2002
11	192	2008	51	61	2001	93	329	2009
12	2	2008	52	268	2012	94	94	2007
13	353	2002	53	34	2003	95	21	2007
14	168	2006	54	99	2010	96	338	2009
15	328	2003	55	170	2000	97	22	2005
16	97	2007	56	128	2002	98	82	2012
17	77	2003	57	131	2003	99	96	2012
18	140	2008	58	87	2007	100	177	2009
19	75	2004	59	170	2012	101	87	2000
20	243	2012	60	25	2004	102	41	2009
21	104	2000	61	359	2003	103	201	2006
22	258	2006	62	7	2010	104	259	2010
23	21	2012	63	264	2008	105	227	2003
24	7	2012	64	362	2000	106	175	2006
25	187	2012	65	361	2012	107	276	2006
26	356	2003	66	337	2008	108	193	2006
27	320	2003	67	210	2007	109	169	2009
28	320	2003	68	180	2011	110	275	2007
29	320	2003	69	195	2008	111	269	2009
30	299	2002	70	345	2007	112	100	2010
31	113	2006	71	327	2012	113	85	2012
32	75	2012	72	153	2005	114	79	2001
33	314	2006	73	331	2009	115	116	2008
34	55	2008	74	345	2007	116	170	2011
35	54	2008	75	216	2008	117	221	2008
36	28	2007	76	285	2009	118	36	2005
37	298	2006	77	192	2011	119	102	2005
38	160	2001	78	215	2006	120	130	2012
39	66	2001	79	29	2011	121	130	2012
40	12	2010	80	68	2012	122	163	2008
			81	41	2001	123	129	2011
			82	158	2010	124	266	2000

125	117	2004	170	137	2006	215	172	2011
126	138	2007	171	217	2012	216	228	2007
127	277	2001	172	118	2009	217	259	2001
128	186	2006	173	182	2003	218	176	2004
129	199	2012	174	61	2007	219	221	2008
130	209	2004	175	84	2003	220	343	2004
131	211	2009	176	91	2000	221	140	2010
132	194	2010	177	169	2000	222	197	2004
133	260	2006	178	8	2005	223	138	2007
134	168	2004	179	187	2003	224	191	2001
135	315	2006	180	12	2000	225	10	2005
136	350	2004	181	65	2012	226	283	2011
137	125	2009	182	178	2010	227	324	2008
138	39	2004	183	9	2007	228	333	2011
139	134	2012	184	283	2002	229	200	2008
140	138	2001	185	281	2000	230	323	2001
141	122	2005	186	193	2007	231	202	2002
142	249	2001	187	94	2000	232	105	2011
143	108	2004	188	129	2011	233	218	2004
144	1	2006	189	160	2000	234	183	2005
145	225	2004	190	254	2010	235	337	2008
146	117	2004	191	147	2009	236	337	2008
147	291	2004	192	229	2003	237	102	2005
148	251	2008	193	10	2005	238	130	2012
149	209	2009	194	262	2011	239	350	2004
150	284	2011	195	130	2003	240	11	2006
151	81	2010	196	301	2007	241	167	2010
152	224	2010	197	131	2012	242	321	2011
153	209	2002	198	225	2002	243	199	2009
154	364	2007	199	17	2011	244	183	2008
155	164	2011	200	97	2008	245	135	2012
156	100	2010	201	183	2005	246	190	2009
157	251	2001	202	65	2012	247	333	2011
158	309	2005	203	194	2012	248	57	2000
159	211	2010	204	345	2007	249	52	2000
160	46	2004	205	114	2008	250	78	2008
161	355	2009	206	192	2001	251	165	2006
162	169	2005	207	231	2002	252	52	2000
163	131	2012	208	212	2002	253	4	2000
164	10	2002	209	212	2002	254	251	2007
165	310	2005	210	57	2002	255	203	2008
166	207	2010	211	57	2002	256	122	2012
167	271	2000	212	57	2002	257	115	2001
168	270	2000	213	342	2007	258	135	2010
169	295	2002	214	220	2011	259	332	2003

260	145	2006	296	277	2001	332	114	2006
261	189	2010	297	140	2002	333	83	2001
262	111	2006	298	311	2007	334	171	2012
263	194	2006	299	13	2001	335	202	2002
264	22	2011	300	337	2002	336	142	2009
265	155	2009	301	210	2003	337	31	2000
266	116	2007	302	205	2012	338	78	2007
267	167	2009	303	258	2011	339	314	2006
268	172	2010	304	202	2003	340	137	2004
269	328	2011	305	150	2012	341	356	2008
270	13	2008	306	13	2007	342	70	2007
271	122	2005	307	13	2007	343	72	2003
272	217	2009	308	116	2003	344	100	2012
273	250	2002	309	263	2011	345	113	2004
274	7	2002	310	332	2012	346	113	2004
275	200	2002	311	23	2002	347	67	2011
276	277	2004	312	334	2010	348	111	2010
277	279	2006	313	264	2011	349	252	2009
278	235	2007	314	99	2009	350	344	2011
279	284	2011	315	73	2005	351	69	2001
280	240	2010	316	284	2006	352	244	2009
281	192	2006	317	295	2006	353	178	2004
282	279	2006	318	100	2000	354	338	2006
283	38	2002	319	202	2007	355	49	2002
284	313	2003	320	217	2005	356	59	2012
285	118	2008	321	10	2012	357	276	2005
286	68	2001	322	113	2009	358	59	2012
287	356	2004	323	36	2005	359	269	2003
288	284	2005	324	157	2011	360	114	2004
289	238	2000	325	84	2002	361	2	2001
290	211	2010	326	36	2001	362	112	2007
291	190	2000	327	87	2003	363	315	2006
292	73	2007	328	10	2012	364	32	2003
293	116	2007	329	221	2002	365	337	2008
294	13	2001	330	63	2011			
295	199	2002	331	1	2009			

11 ANEXO II

